UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

Deterministics, Initial Conditions and Breaks in Long Memory Time Series

Autor:
Heiko Rachinger

Director:
Carlos Velasco Gómez

DEPARTAMENTO DE ECONOMÍA

Getafe, Junio 2012

TESIS DOCTORAL

DETERMINISTICS, INITIAL CONDITIONS AND BREAKS IN LONG MEMORY TIME SERIES

Autor: Heiko Rachinger

Director/es: Carlos Velasco Gómez

Firma del Tribunal Calificador:

Firma

Presidente:

Vocal:

Vocal:

Vocal:

Secretario:

Calificación:

Getafe, de de

Acknowledgements

En primer lugar, muchísimas gracias a mi director de tesis Carlos, quien, con su ayuda, apoyo y paciencia durante todos estos años, ha hecho posible esta tesis.

Many thanks to Laura Mayoral for her very helpful comments on the first version of the thesis and to Jesús Gonzalo and Uwe Hassler for their suggestions in addition to several helpful discussions over the last years. I am also indebted to Miguel Delgado, Peter M. Robinson and Abderrahim Taamouti for having read some chapters of the thesis. I am especially grateful to Juanjo Dolado for all the support and guidance over the last 5 years.

Thanks to the London School of Economics for the hospitality and Javier Hidalgo for the help during my stay in Spring 2010. I also thank participants of the ENTER Jamboree 2010, the Econometric Society World Congress 2010, the Simposio de la Asociación Española de Economía 2010, ERCIM 2010 ESEM 2011, SAEe 2011,SNDE 2012 and III$_t$ Workshop in Time Series Econmetrics. I further thank participants of seminars at London School of Economics, Universidad Carlos III and Universidad Islas Baleares. I am especially grateful to all researchers that discussed this thesis during my job market seminars. I am highly indebted to Henghsiu Tsai for always believing in me. I further, thank Academia Sinica for the hospitality during my stay in Spring 2011. I acknowledge financial support from the Spanish Ministerio de Educación y Ciencia, Ref. no. SEJ2007-62908/ECON.

Many friends have helped me stay sane through these difficult years. Thank you very much, Ming-jin, for thousands of unforgettable moments over the last years and for always being there when I need you. Muchas gracias a Román por haber sido un compañero de oficina bárbaro (a pesar del frio terrible!) y un todavía mejor amigo. Muchas gracias también por miles de mates buenísimos. A Efi por haber sido una muy buena amiga durante todos estos años y por siempre creer en mí. A Eva por haber escuchado mis quejas eternas durante todo el tiempo y por toda la ayuda. Muchísimas gracias a Vanessa no solo por ser mi referencia académica sino también y más importante por ser una muy buena amiga. Gracías también a Joaquín por haber sido un gran compañero y amigo. A Miguel por eternas discusiones futboleras. A Barbara por creer en mí durante mi modesta carera académica desde Maastricht. A María por todos los lunch breaks y tu amistad. A Noelia por ser siempre tan simpática. A Ana por los coffee breaks y charlas. Ánimo con todo! Thanks to Georgi for the time we shared, filled with weights, beers and enriching conversations. Thanks to Dani for having shared all the Hare Krishna meals in London. I also want to thank all the other PhD friends who have been with me along the way: Wang, Yunrong, María, Luis...

Thank you very much for many nice moments in Getafe and elsewhere. I deeply hope to keep in touch. Be it, in journeys to and from Vienna or be it in conferences along our carreers, promising for some, not so for others.

Also many thanks to my friends at home, especially Örni and Martin, for their unconditional friendship.

Most importantly, I want to thank my family, my parents Karl and Erna, my sister Kerstin and my brothers Bernd and Alexander, for their unconditional support and patience. To them I dedicate this thesis. Thank you for everything!!!

Resumen en Castellano

En mi tesis doctoral, se modelizan series temporales con memoria larga y con un componente determinista que potencialmente sufre rupturas. Se consideran contrastes para rupturas y la estimación de los parámetros. Finalmente, se analiza la estimación eficiente de tendencias lineales y su impacto proveniente de la presencia y la longitud de de la pre-muestra.

En el primer capítulo, **"Multiple Breaks in Long Memory Time Series"**, se propone un enfoque unificado para la modelización de rupturas en la memoria y la media de una serie temporal. Las series temporales macroeconómicas y financieras a menudo muestran características de memoria larga, como funciones de auto-correlación que decaen hiperbólicamente. Ha habido una larga discusión sobre si tales series temporales se pueden describir por modelos fraccionalmente integrados o si la memoria larga es espuria debido a rupturas en su media. Si bien el número de rupturas es conocido, la fracción de ruptura y los parámetros en los diferentes regímenes se estiman conjuntamente por el método de mínimos cuadrados ordinarios no lineales. El estimador de la fracción de ruptura resulta ser súper-consistente, con una tasa T tanto para rupturas en la memoria como en la media. Por otra parte, se analizan contrastes F para determinar el número de rupturas cuando este número es desconocido. Su comportamiento asintótico depende de funcionales de movimientos Brownianos estándares y fraccionales. Puesto que una ruptura en la media provoca un rechazo espurio del contraste para la ruptura en la memoria, es difícil identificar qué parámetro está cambiando en cada punto de ruptura encontrado. Para resolver este problema, se propone un procedimiento secuencial, como instrumento robusto, con el fin de detectar rupturas en cada parámetro sin causar efectos espurios que pudieran sugerir rupturas en el otro parámetro. Para mejorar el comportamiento en muestras pequeñas, se propone utilizar contrastes basados en el método "Bootstrap", para los que se deriva la validez y la consistencia. Como ejemplo ilustrativo, se aplica esta metodología para el análisis de una serie mensual de la inflación estadounidense.

En el segundo capitulo, **"Lagrange Multiplier and Wald Tests for Breaks in the Memory and the Level of a Time Series"** (*con Juan J. Dolado y Carlos Velasco*), se analizan contrastes de multiplicador de Lagrange (LM) y Wald para captar la presencia e identificar el número de rupturas en la memoria y en el nivel de una serie temporal. Por un lado, el contraste LM tiene la ventaja de que la estadística del contraste se deriva bajo la hipótesis nula y por tanto permite alternativas más generales. Por otro lado, el contraste de Wald explota informaciones adicionales sobre la hipótesis alternativa, y por lo tanto tiene mayor potencia en comparación con el LM. En este capítulo se analizan tanto los casos de fracciones de ruptura conocidas y desconocidas. Se derivan las distribuciones asintóticas y se muestra que estos contrastes tienen la misma distribución asintótica que el contraste F del primer capítulo tanto bajo la hipótesis nula como bajo una alternativa local general. Se generalizan los contrastes permitiendo una dinámica a corto plazo potencialmente con rupturas, además de cuantificar el efecto sobre la potencia local proveniente de la estimación de los parámetros en el primer régimen. Para comparar los contrastes más en detalle, comparamos la potencia bajo la alternativa mostrando que para rupturas en la memoria y en el nivel, el contraste Wald domina al contraste F y éste último al contraste LM.

En el tercer capítulo, **"Linear Trends, Fractional Trends and Initial Conditions"**, se analiza la estimación eficiente de la tendencia lineal de una serie temporal. Las series temporales macroeconómicas a menudo se caracterizan por seguir tendencias que se pueden describir de forma lineal. La presencia de ruido aditivo de memoria larga, anidando como caso particular la situación de una raíz unitaria, y el supuesto de la existencia y longitud de la historia pre-muestral tienen un impacto en la estimación eficiente de la tendencia lineal. Definiendo la condición inicial como el puente entre dos definiciones diferentes de memoria larga – una historia pre-muestral inexistente e infinita–, se compara el comportamiento asintótico de diferentes estimadores de la tendencia y se discute su eficiencia como función de la longitud de la condición inicial. Para el caso de la historia pre-muestral inexistente un estimador de mínimos cuadrados generalizados que corrige la estructura de la dependencia específica puede aportar beneficios enormes en términos de eficiencia. Sin embargo, para condiciones iniciales ligeramente más remotas, este estimador pierde su eficiencia. Así, es la presencia en lugar de la longitud de la condición inicial lo que importa en la elección del estimador más eficiente para la tendencia lineal. Por lo tanto, este trabajo complementa la literatura existente sobre la estimación eficiente de las tendencias lineales en el caso de historia pre-muestral infinita. Para ilustrar la metodología, se estiman las tasas de crecimiento del producto interior bruto de tres países y se contrasta si estas tasas son positivas.

Dissertation Abstract

In my thesis, **"Deterministics, Initial Conditions and Breaks in Long Memory Time Series"**, I model long memory time series with a potentially breaking deterministic component. I consider testing for breaks as well as parameter estimation. Further, I analyze efficient estimation of linear time trends and the impact from the presence and length of the pre-sample.

In the first chapter, **"Multiple Breaks in Long Memory Time Series"**, I propose a unified approach for modeling breaks in the memory and in the mean of a time series. Macroeconomic and financial time series often display long memory characteristics such as hyperbolically decaying autocorrelation functions. There has been a long discussion whether these time series can be described by fractionally integrated models or whether the long memory is spurious due to breaks in their mean. If the number of breaks is known, the break fraction and the parameters in the different regimes are jointly estimated by least squares. The break fraction estimator is superconsistent at rate T for breaks in the memory and/or mean. Furthermore, I analyze F-tests for determining the number of breaks, if this number is unknown. Their asymptotic behavior depends on functionals of standard and fractional Brownian Motions. Since a break in the mean causes a spurious rejection of the test for a break in the memory, it is difficult to identify which parameter is breaking at each break point. To solve this problem, I propose a sequential procedure as a robust instrument to detect breaks in each parameter without spurious effects from breaks in the other parameter. To improve the finite sample behavior, I suggest using bootstrap based tests for which I derive validity and consistency. Finally, I apply the methodology to analyze the monthly US inflation series.

In the second chapter, **"Lagrange Multiplier and Wald Tests for Breaks in the Memory and the Level of a Time Series"** (*with Juan J. Dolado and Carlos Velasco*), we analyze Lagrange Multiplier (LM) and Wald tests for the presence and number of breaks in memory and level of a time series. The advantage of LM tests is that they are derived under the null hypothesis, thus, allowing for more general alternative hypotheses. On the other hand, the Wald test can exploit further information on the alternative, potentially leading to higher power. We analyze the cases of known and unknown break points. We derive the asymptotic distributions and show that these tests have the same asymptotic distribution as the F-test in the first chapter both under the null and under a local alternative. Further, we

extend the proposed testing procedure by allowing for potentially breaking short run dynamics. We quantify the effect coming from the estimation of the parameters in the first regime. In order to compare the tests in more detail, we compare their power under the alternative and show that for both breaks in the memory and the level, the Wald test dominates the F-test and this dominates the LM test.

In the third chapter, **"Linear Trends, Fractional Trends and Initial Conditions"**, I analyze the efficient estimation of a linear trend. Macroeconomic time series further often have a trending behavior that can be described by models including a linear time trend. The presence of additive long memory noise, nesting as a special case the unit root situation, and the assumption about existence and length of the pre-sample history have an impact on the efficient trend estimation. Defining the initial condition as the bridge between two alternative definitions of Long Memory – zero and infinite pre-sample history –, I compare the asymptotic behavior of different trend estimators and discuss their efficiency as a function of the length of the initial condition. For the case of no pre-sample history, a generalized least squares estimator that corrects the specific dependence structure in this case brings huge efficiency gains. However, for already slightly more remote initial conditions, this estimator loses its efficiency. Thus, the presence rather than the length of the initial condition matters for the choice of the best trend estimator. Thus, I complement the existing literature on efficient trend estimation in the infinite pre-sample case. In order to illustrate the methodology I estimate the growth rates of three countries and test whether these rates are positive.

Contents

Chapter 1

Multiple Breaks in Long Memory Time Series

Abstract. We analyze least squares (LS) estimation of breaks in long memory time series. We show that the estimator of the break fraction is consistent and converges at rate T when there is a break in the mean, in the memory or in both parameters. Further, we analyze tests for the number of breaks. When testing for breaks in the memory, the asymptotic results correspond to standard ones in the literature. When testing for breaks in the mean and when testing for breaks in both parameters, the results differ in terms of the asymptotic distribution of the test statistic. In this case, the LS-procedure loses its asymptotic pivotality. We further propose a method in order to distinguish between long memory, breaks in the memory and breaks in the mean. Such a distinction is difficult but is important for reasons such as shock identification, forecasting and detection of spurious fractional cointegration. In a simulation exercise, we find that the tests based on asymptotic critical values are oversized in finite samples. Therefore, we suggest using the bootstrap, for which we derive validity and consistency, and we confirm its better size properties. Finally, we use the method to test for breaks in the U.S. inflation rate.

1.1 Introduction

Macroeconomic and financial time series are in general persistent and display long memory characteristics such as hyperbolically decaying autocorrelation functions (see e.g. Ding et al., 1993, and Gil-Alana and Robinson, 1997). There has been a long discussion whether these time series can be described as fractionally integrated models or whether their long memory is spurious due to breaks in their mean (see e.g. Lobato and Savin, 1998, Granger and Hyung, 2004). Recently, Perron and Qu (2010) discuss that many time series are more likely generated by stationary processes with a break in their mean rather than by long memory models. However, processes with breaks in the long memory parameter can also generate those series (McCloskey, 2010). Thus, it is hard to distinguish between long memory, breaks in the memory and breaks in the mean. However, such a distinction is important in practice for reasons such as shock identification, detection of spurious cointegration, forecasting and economic modelling.

In Chapter 3 of this thesis, we further motivate the use of a FI(d) model and the joint modelling of breaks in level and in memory by considering the time series properties of inflation rate. It is known that long memory can arise from aggregating individual series (see e.g. Zaffaroni, 2004). This example connects changes in the parameters of individuals firms with changes in the memory and level of the aggregate prices. In particular, a change in the degree of competition in the product and services markets implies changes in persistence and level, changes in the monetary policy imply a sole change in the level. Finally, changes in both parameters could only cause a change in the memory.

The aim of this paper is to provide a method to detect the presence of breaks in memory and in mean and to distinguish between them. We propose a unified approach for modeling breaks in the mean and the memory. In particular, we extend the Bai and Perron (1998) methodology to the long memory context and analyze least squares estimation of breaks in long memory time series. In their short memory framework, they discuss a linear model with multiple breaks. They derive consistency and $T-$rate convergence of the break fraction estimate and the asymptotic distribution of the parameter estimates in the regimes. Finally, they provide a series of tests for the existence and number of breaks. Boldea and Hall (2010) extend Bai and Perron's (1998) analysis into a nonlinear setting. They show that the results of Bai and Perron (1998) do not change, even though the proofs become more involved. By considering nonlinear models, they encompass several ergodic models but not long memory time series models.

Kuan and Hsu (1998) and Lavielle and Moulines (2000) analyze the LS procedure for a process with a break in the mean and a stationary long memory error term, yet without breaks in the memory. Since they do not integrate explicitly the memory parameter in their analysis, they find different asymptotics. Shao (2011) proposes a test for a break in the mean under long range dependence, extending a test proposed in Hidalgo and Robinson (1996) towards allowing for an unknown break fraction. Further, Gil-Alana (2008) analyzes a similar methodology as ours. Nevertheless, he works with a data generating process that is not a typical long memory process. He also does not derive rigorously the asymptotic distributions of the estimates and statistics. He conjectures that the asymptotic properties resemble the ones found in Bai and Perron (1998). However, we show that the critical values employed in Gil-Alana (2008) are not the correct ones for testing for breaks in the mean. Besides, Gil-Alana (2008) is not specific about the impact coming from the estimation of the memory parameter d. Taking the latter into account, the problem becomes a nonlinear one and we have to consider specific arguments to derive the asymptotic properties. Yamaguchi (2011) analyzes a parametric estimator for the break fraction when there are breaks in the long memory parameter. However, he implicitly uses the same data generating as Gil-Alana (2008) does. Further, he assumes a *zero* mean in his analysis. Therefore, his analysis also differs from ours.

In this paper, we derive consistency and T-rate convergence of the break fraction estimator and the asymptotic distribution of the parameter estimates when there are breaks in the memory and/or the mean. We assess the power of break tests by considering local breaks in the memory and in the mean. The asymptotic distribution of these tests differ from the ones of Bai and Perron (1998) and the procedure loses its asymptotic pivotality. We discuss tests for determining which parameter is the changing one. Since the tests based on asymptotic critical values suffer from some size distortions in finite samples, we suggest using the bootstrap for which we derive validity and consistency.

Another strand of literature focuses on testing for the presence and the number of breaks in the memory parameter in time series with long memory. Beran and Terrin (1996) use parametric Whittle estimators to test for a break in the memory. Hassler and Meller (2011) introduce an augmented Lagrange Multiplier test to test semiparametrically for breaks in the memory, allowing for breaks in the mean. Hassler and Scheithauer (2011) show that tests for the null hypothesis of I(0) series against alternatives of a change from I(0) to I(1), discussed by Kim, Belaire-Franch and Amador (2002) and Busetti and Taylor (2004), are also consistent for a change from I(0) to I(d), for $d > 0$. Sibbertsen and Kruse (2009) derive a CUSUM of squares-based test. Sibbertsen and Willert (2009) show that these tests are sensitive to breaks in the mean and simulate critical values that are valid in the presence of mean shifts. Martins and Rodrigues (2010) use recursive forward and backward estimation of a LM test. McCloskey (2010) uses a modified ratio of weighted partial sums to test semiparametrically for breaks in the memory. Finally, Lavancier et al.

(2011) modify the procedure by Kim et al. (2002) to improve the power of the test. They further, distinguish between a gradual and abrupt change in the memory, the latter corresponding to the data generating process that is analyzed in our paper.

In Section 2, we discuss the model and the least squares estimation of an unstable process. In Section 3, we derive the asymptotic behavior of the estimators in the presence of breaks. In Section 4, we analyze tests for the number of breaks and examine the behavior of these tests in finite samples. In Section 5, we provide a monte carlo analysis of the tests. In Section 6, we propose a sequential testing strategy to determine which parameter is changing. In Section 7, we analyze the bootstrap. In Section 8, we apply the methodology to the U.S. inflation series and test for breaks in memory and mean in this series. Finally in Section 9, we conclude. Some Lemmata and additional Propositions which are needed for the analysis are provided in Appendix A. The proofs are collected in Appendix B.

1.2 Preliminaries

We consider the following model with m breaks in $(T_1^0, T_2^0,..., T_m^0)$ $(m+1$ regimes),

$$y_t = \mu_j^0 + \Delta_t^{-d_j^0} u_t,\ t = T_{j-1}^0 + 1, ..., T_j^0,\ j = 1, ..., m+1. \tag{1.1}$$

The coefficients of interest $\theta_j^0 = (\mu_j^0, d_j^0)$ lie in some set $\Theta_j = M_j \times D_j$. The process consists of an intercept and a Type II fractionally integrated disturbance,

$$\Delta_t^{-d_j^0} u_t = \sum_{k=0}^{t-1} \pi_k(-d_j^0) u_{t-k}, \tag{1.2}$$

where Δ_t^{-d} denotes the truncated fractional differencing filter with memory d and where

$$\pi_k(-d) = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}, k = 0, ..., t-1,$$

denotes the sequence of coefficients of the expansion of Δ_t^{-d}. In this and in the next section, we assume that the number of breaks, m, is known but the actual break points, $(T_1^0, T_2^0, ..., T_m^0)$, are unknown. The latter will be estimated together with the parameter vector $(\theta_j^0)_{j=1}^{m+1}$. We consider equally the cases of a pure structural change model, in which both coefficients change, and a partial structural change model, in which some coefficient does not change.

For obtaining the conditional sum of squares (CSS) estimator in a stable context, it suffices to apply the filter Δ_t^d to the process since for $d = d^0$, the resulting residuals are u_t. Nevertheless, for the unstable process (1.1), it is not correct to apply the filter $\Delta_t^{d_j^0}$ to the process (1.1), as it is done in Gil-Alana (2008) or in Yamaguchi (2011), because $\Delta_t^{d_j^0} y_t$ is a weighted sum of $I(d_1)$ to $I(d_j)$ terms rather than u_t. In order to avoid this problem, Dolado *et al.* (2009) define the process implicitly and in Chapter 3, we define it explicitly as

$$\Delta_t^{d_j^0}\left(y_t-\mu_j^0\right)=u_t,\ t=T_{j-1}^0+1,...,T_j^0. \tag{1.3}$$

In this case it suffices to apply the fractional differencing filter $\Delta_t^{d_j^0}$ to obtain $I(0)$ residuals and the whole analysis simplifies considerably. The transition of the memory and the mean is smooth. Given the persistent nature of the process, this transition can occur rather slowly for the mean. Thus, the parameter μ_j in (1.3) does not correspond to the mean in regime j. Notice that Lavancier et al. (2011) discuss an alternative gradual transition of the memory. However, the process defined in (1.3) is not strictly a $I(d_j^0)$ process in $t>T_1^0$. Therefore, we rather apply a filter to data generated by (1.1) that restricts the filtered data to lie in the interval of the corresponding regime. First, we define a break fraction λ_i and the true break fraction λ_i^0 as T_i/T and T_i^0/T respectively. In particular, we set the residuals

$$\hat{u}_t\left(\lambda_{j-1},\theta_j\right)=\Delta_{t-[\lambda_{j-1}T]}^{d_j}\left(y_t-\mu_j\right),\ t=T_{j-1}+1,...,T_j. \tag{1.4}$$

Since the fractional differencing filter for regime j is restricted to the observations of this regime, this filter avoids the aforementioned mixing of observations from different regimes. The resulting residuals in (1.4) are close to $I(0)$, if break fraction and coefficients are estimated close to the true ones. However, apart from terms coming from the distance between estimate and true break fraction and coefficients, there are also some additional *initial condition* terms coming from the fact that the applied fractional filter is too short. These terms are similar in nature to the *initial condition* terms that show up when applying a truncated Type II fractional filter to an untruncated Type I process. The technical difficulties arise from showing that all these terms are asymptotically negligible.

In particular, assume the process has m breaks at $(T_1^0,...,T_m^0)$, where the true number of breaks m is known. We estimate the break fractions $\lambda_j=T_j/T$ together with the coefficients in the regimes by conditional sum of squares (CSS) estimation. Let

$$S_T\left(\boldsymbol{\lambda},\boldsymbol{\theta}\right)=\sum_{i=1}^{m+1}S_{i,T}\left(\lambda_{i-1},\lambda_i,\theta_i\right)=\sum_{i=1}^{m+1}\sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_iT]}\hat{u}_t\left(\lambda_{i-1},\theta_i\right)^2, \tag{1.5}$$

where $\hat{u}_t$ is defined in (1.4). For simplicity, we illustrate the procedure for $m=1$, the general case follows equally. For a given break fraction λ_1 with $T_1=[\lambda_1T]$ and

(d_1, d_2),

$$\{\hat{\mu}_i(d_i, \lambda_1)\}_{i=1,2} = \underset{\mu_1,\mu_2 \in M_1 \times M_2}{\operatorname{argmin}} \{S_{1,T}(0, \lambda_1, \mu_1, d_1) + S_{2,T}(\lambda_1, 1, \mu_2, d_2)\}.$$

Substituting the estimator $\{\hat{\mu}_i(d_i, \lambda_1)\}_{i=1,2}$ into the objective function, we obtain the conditional memory estimator

$$\{\hat{d}_i(\lambda_1)\}_{i=1,2} = \underset{d_1,d_2 \in D_1 \times D_2}{\operatorname{argmin}} \{S_{1,T}(0, \lambda_1, \hat{\mu}_1(d_1), d_1) + S_{2,T}(\lambda_1, 1, \hat{\mu}_2(d_2), d_2)\}.$$

Finally, we minimize the objective function with respect to λ_1 and obtain an estimator for the break fraction as

$$\hat{\lambda}_1 = \underset{\lambda_1 \in [\epsilon, 1-\epsilon]}{\arg\min} S_{1,T}\left(0, \lambda_1, \hat{\mu}_1\left(\hat{d}_1(\lambda_1), \lambda_1\right), \hat{d}_1(\lambda_1)\right) + S_{2,T}\left(\lambda_1, 1, \hat{\mu}_2\left(\hat{d}_2(\lambda_1), \lambda_1\right), \hat{d}_2(\lambda_1)\right).$$

The estimators for the parameters d_i and μ_i $(i = 1, 2)$ are

$$\hat{d}_i(\hat{\lambda}_1) \text{ and } \hat{\mu}_i(\hat{d}_i(\hat{\lambda}_1), \hat{\lambda}_1).$$

The truncated filter (1.4) is attractive because it estimates the parameters in the different regimes separately. Therefore, considering m breaks is conceptionally not more involved than considering *one* break. Besides, it extends easily to a Type I process DGP, $\Delta_\infty^{-d_i^0} u_t$. The only difference is that for a Type I process, the truncated part is $\sum_{j=0}^{t-1} \pi_j(d) \Delta_\infty^{-d_i^0} u_{t-j}$ rather than $\sum_{j=1}^{t} \pi_j(d) \Delta_{t-j}^{-d_i^0} u_{t-j}$.

For the subsequent analysis we need the following assumptions:

Assumption 1.

(i) The error term u_t is $iid\ (0, \sigma^2)$.

(ii) $E|u_t|^s < \infty$, $s > \frac{3}{2\left(1-2\max\left(d_i^0\right)\right)}$.

Assumption 2. The common parameter space is $\Theta = M \times D = ([\underline{\mu}, \bar{\mu}], [0, 1/2 - \varepsilon])$, $0 < \varepsilon < 1/2$, and $\theta^0 \in \Theta$.

Assumption 3. $T_i^0 = \left[T\lambda_i^0\right]$, $i = 1, ..., m$, where $0 < \lambda_1^0 < ... < \lambda_m^0 < 1$.

Assumption 1 implies that the errors are independent from the regression function $f_t(\theta) = (\Delta_{t-T_{i-1}}^{d_i} - 1)(y_t - \mu_i)$, $E[u_t f_t(\theta)] = 0$ for all θ and t. In contrast to Boldea and Hall (2010), our regressor is not strictly stationary $\alpha - mixing$ but fractionally integrated. For further generalizations of the error term, we could assume a different variance in the different regimes or a short memory error process, $u_t = w_\theta(L)\varepsilon_t$. In the former, for $m = 1$, let $u_t^{(1)}$ and $u_t^{(2)}$ denote the errors of the two regimes. The variance of the mean estimator of the second regime depends then on both error variances. For the latter, the analysis is complicated by the correlation between the estimators of $w_\theta(L)$ and d. Hualde and Robinson (2011) analyze the case of this estimator in a stable context with short term component but without mean. In the following sections, we also consider the case of a stable autoregressive structure. Further, we discuss briefly the case of testing for a changing short term component and conjecture that the asymptotic distributions follow from combining Boldea and Hall's (2010) approach with ours. Assumption 1 Part (ii) is needed for weak convergence of partial sums of products of the regressor and the error term. Assumption 3 is a standard assumption in the break literature.

For the following analysis of the estimators in the presence of structural breaks, we need to analyze the behavior of the CSS estimator of one parameter if the other one is not consistently estimated. For simplicity, we illustrate the problem for the stable case. First, the CSS estimator of the memory works well when there is no deterministic component, or when it is known or consistently estimated at rate $T^{1/2-d^0}$. On the other hand, if the mean is not consistently estimated, the memory estimator can have a huge bias in finite samples (Chung and Baillie, 1993). But there are no asymptotic results for this case to my best knowledge. Proposition 1a) delivers these results. Equally, we analyze the properties of the mean estimation when the memory is inconsistently estimated. Proposition 1b) shows that consistency and rate of convergence of the mean estimation are asymptotically not affected by the memory estimation.

Proposition 1 *(Behavior of the CSS estimator)*
a) Given $\mu \in Int(M)$ and $d^0 \in Int(D)$,

$$\hat{d}(\mu) - d^0 = O_p(T^{-1/2}) \text{ uniformly in } \mu.$$

b) Given $d \in D$,

$$\hat{\mu}(d) - \mu^0 = O_p(T^{d^0-1/2}) \text{ uniformly in } d.$$

Therefore, if the mean is inconsistently estimated or not estimated – $\hat{\mu} = 0$ –, the estimation of the memory is inconsistent if $d^0 = 0$ but still consistent if $d^0 \in Int(D)$. The finite sample effects depend on $d^0, (\mu^0 - \mu)$ and T. Especially, for d^0 close to 0, the estimate can be highly upward biased in finite samples.

In the following sections, we analyze long memory time series with a break only in the mean μ, only in the memory d or in both parameters.

1.3 Asymptotic behavior of estimates in the presence of breaks

Given the nonlinear nature of our problem, our approach is closer to Boldea and Hall (2010) rather than to Bai and Perron (1998). However, our process is fractionally integrated and does not meet their conditions. In the following, we derive most of the results newly.

The break fraction estimate is consistent for breaks in the memory, in the mean and in both parameters.

Theorem 2 *(Consistency of the break fraction estimator)*
Under Assumptions 1-3,

$$\hat{\lambda}_i \xrightarrow{p} \lambda_i^0.$$

Using consistency of the break fraction estimates, we establish their rate of convergence.

Theorem 3 *(Rate of convergence of the break fraction estimator)*
For every $\eta > 0$, there exists a finite $C > 0$ such that for all large T,

$$P\left(T|\hat{\lambda}_i - \lambda_i^0| > C\right) < \eta.$$

We find a $T-$rate convergence for the break fraction estimator when there are breaks in the memory, in the mean or in both parameters. This $T-$rate corresponds to the one found in Lavielle and Moulines (2000) for a break in the mean in a process with Type I long memory error but is faster than the one found in Kuan and Hsu (1998). Given the T rate convergence of the break fraction estimates, Theorem 4 provides consistency, the rate of convergence and the limiting distribution of the parameter estimates. The estimators $\hat{d}_i$ and $\hat{d}_j$ are asymptotically uncorrelated and the estimators $\hat{\mu}_i$ and $\hat{\mu}_j$ are correlated.

Theorem 4 *(Asymptotic distribution of the CSS estimators)*
Under Assumptions 1-3, with $\theta^0 \in Int(\Theta)$, $i = 1, ..., m$,

$$diag\left(T^{1/2}, T^{1/2-d_i^0}\right)\left(\hat{\theta}_i - \theta_i^0\right) \xrightarrow{d} N\left(\underline{0}, D_i\left(d_i^0, \lambda_i^0, \lambda_{i-1}^0\right)\right),$$

where

$$D_i\left(\lambda_{i-1}^0, \lambda_i^0, d_i^0\right) = \begin{pmatrix} \frac{6}{\pi^2}\left(\lambda_i^0 - \lambda_{i-1}^0\right)^{-1} & 0 \\ 0 & \sigma^2\left(\frac{\Gamma^2\left(1-d_i^0\right)\left(1-2d_i^0\right)}{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}} + D_{ii}^{\mu}\left(\lambda_{i-1}^0, \lambda_i^0, d_i^0\right)\right) \end{pmatrix},$$

where $\hat{\theta}_i$ and $\hat{\theta}_j$ are jointly normal, $\hat{d}_i$ and $\hat{\mu}_j$ are asymptotically uncorrelated for $i, j = 1, 2$, and $\hat{d}_i$ and $\hat{d}_j$ are uncorrelated and $\hat{\mu}_i$ and $\hat{\mu}_j$ are correlated for $i \neq j$ with a covariance $\sigma^2 D^{\mu}_{ij}(\{\lambda^0_{k-1}, \lambda^0_k, d^0_k\}_{k=i,j})$.

$D^{\mu}_{ii}\left(\lambda^0_{i-1}, \lambda^0_i, d^0_i\right)$ is the variance component arising from applying the too short differencing filter on the fractionally integrated error series

$$D^{\mu}_{ii}\left(\lambda^0_{i-1}, \lambda^0_i, d^0_i\right) = \frac{\Gamma^4\left(1-d^0_i\right)\left(1-2d^0_i\right)^2}{\left(\lambda^0_i - \lambda^0_{i-1}\right)^{2-4d^0_i}} A^{\mu}_i\left(\lambda^0_{i-1}, \lambda^0_i, d^0_i\right), \tag{1.6}$$

where

$$\begin{aligned} & A^{\mu}_i\left(\lambda^0_{i-1}, \lambda^0_i, d^0_i\right) \\ = & \lim_{T\to\infty} T^{-1} \sum_{k=1}^{[\lambda^0_{i-1}T]} \left(T^{d^0_i} \sum_{t=1}^{[(\lambda^0_i - \lambda^0_{i-1})T]} \pi_{t-1}\left(d^0_i - 1\right) \sum_{l=0}^{t} \pi_l\left(d^0_i\right) \pi_{[\lambda^0_{i-1}T]+t-l-k}\left(-d^0_i\right) \right)^2 . \end{aligned} \tag{1.7}$$

$D^{\mu}_{ij}(\{\lambda^0_{k-1}, \lambda^0_k, d^0_k\}_{k=i,j})$ is defined as (1.35) in the Appendix. Both are functions of $\{\lambda^0_{k-1}, \lambda^0_k, d^0_k\}_{k=i,j}$ and have to be numerically approximated. We estimate the covariance matrix of the estimator by replacing $\{\lambda^0_{i-1}, \lambda^0_i, d^0_i\}$, D^{μ}_{ii} and D^{μ}_{ij} by their estimates and $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} \hat{u}^2_t$.

Finally, if there are some short run dynamics in the form of a stable and known causal autoregressive structure of order p (AR(p)),

$$\alpha\left(L\right)\left(y_t - \mu^0_i\right) = \Delta_t^{-d^0_i} \varepsilon_t, \; T^0_{i-1} < t \leq T^0_i, \tag{1.8}$$

the mean estimation behaves as in Theorem 4. The memory estimator is correlated with the estimator of the AR component. In particular,

$$Var\left(T^{1/2}(\hat{d}_i - d^0_i)\right) = \omega^{-2}\left(\lambda^0_i - \lambda^0_{i-1}\right)^{-1},$$

where $\omega^2 = \frac{\pi^2}{6} - \kappa'\Phi^{-1}\kappa$, $\kappa = (\kappa_1, ..., \kappa_p)'$ and $\kappa_k = \sum_{j=k}^{\infty} j^{-1} c_{j-k}, k = 1, ..., p$ where c_j are the coefficients of L^j in the expansion of $1/\alpha\left(L\right)$. $\Phi = \left[\Phi_{k,j}\right], \Phi_{k,j} = \sum_{t=0}^{\infty} c_t c_{t+|k-j|}, k, j = 1, ..., p$ denotes the Fisher information matrix for α under Gaussianity. The proof follows from combining Hualde and Robinson (2011) and our Theorem 4.

1.4 Tests

Up to now, we have assumed that the number of breaks is known. In the following, we analyze tests for determining the number of breaks if this number m is unknown.

1.4.1 F-test of 0 versus k breaks

First, we consider the hypothesis of no breaks and the alternative of k breaks, where in practice k is a small number:

$$H_0 : m = 0 \text{ vs. } H_1 : m = k.$$

Let $\boldsymbol{\lambda}$ denote a break fraction partition satisfying the standard assumption of asymptotic distinctiveness and distance to the end-points. In particular, $\boldsymbol{\lambda}$ belongs to the subset

$$\Lambda_\epsilon = \{\boldsymbol{\lambda} \equiv (\lambda_1, ..., \lambda_k) : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_i \geq \epsilon, \lambda_i \leq 1 - \epsilon\}$$

with $\epsilon > 0$. Given a break partition $\boldsymbol{\lambda}$, let

$$SSR_k(\boldsymbol{\lambda}) = \min_{\theta_1,...,\theta_{k+1}} \sum_{i=1}^{k+1} \sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \left(\Delta_t^{d_i}(y_t - \mu_i)\right)^2 \tag{1.9}$$

denote the minimized sum of squared residuals under the alternative hypothesis of k breaks. Note that this filter differs from the previous filter (1.4) in being truncated at 1 rather than at $[\lambda_{i-1}T]$. This filter is the appropriate one under H_0. In consequence, also the test statistic will be constructed under the assumption that H_0 is true. From (1.9), we obtain the unconstrained estimators in the $k+1$ regimes, $(\hat{\theta}_1, ..., \hat{\theta}_{k+1})$, given the break partition $\boldsymbol{\lambda}$. Equally, SSR_0 denotes the minimized sum of squares under the hypothesis of no breaks. As in Bai and Perron (1998) and Boldea and Hall (2010), we use a sup F-type test

$$\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon} F_T^{\vartheta}(\boldsymbol{\lambda}, k; p) = \sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon} \frac{(SSR_0 - SSR_k(\boldsymbol{\lambda}))/kp}{SSR_k(\boldsymbol{\lambda})/[T-(k+1)p]}. \tag{1.10}$$

The number of changing parameters p is *one* or *two*. The superscript $\vartheta \in \{d, \mu, (d, \mu)\}$ denotes the parameter in which we are testing for breaks. ϵ is a fixed small number. The larger ϵ is, the larger is the power, but the test might become inconsistent, if Λ_ϵ does not contain the true break fraction under the alternative. For the break only in the memory (mean), $SSR_k(\boldsymbol{\lambda})$ constraints the mean (memory) to be constant over the regimes.

Since from (1.9), the same μ_i is subtracted from observations with true mean μ_j^0 of all regimes $j \leq i$, the mean μ_i^0, $i > 1$, is inconsistently estimated under the alternative hypothesis. This does not happen for the memory estimator d_i since the terms arising from applying the wrong filter are negligible. Alternatively, the filter (1.4) from Sections 2 and 3 would solve this problem of inconsistent estimation under the alternative. However, for determining the asymptotic distribution of $\sup_{\boldsymbol{\lambda}} F_T$ under H_0, the filter in expression (1.9) is more appropriate. The asymptotic distribution resembles the one of Bai and Perron (1998) and the size properties are better than the ones with filter (1.4). Despite the estimators are inconsistent, this test has nontrivial power against local alternatives (Theorem 5) and is consistent (Theorem 7).

We consider the following local alternative for assessing the power of the tests for processes close to H_0,

$$H_{1,T} : d_t^0 = d_1^0 + T^{-1/2} h_d \left(\frac{t}{T}\right) \text{ and } \mu_t^0 = \mu_1^0 + T^{d_1^0 - 1/2} h_\mu \left(\frac{t}{T}\right).$$

As in Lazarová (2005), $h_j(\frac{t}{T})$, $j = d, \mu$, is a bounded variation function on [0,1]. This local alternative comprises many types of structural change models. A function $h(\tau) = \sum_{j=1}^{i} \delta_j I\left(\lambda_j^0 \leq \tau\right)$ describes abrupt breaks of size δ_i at time $[\lambda_i^0 T]$. A function h consisting of constant segments connected by smooth curves describes a smooth transition between the different levels of the parameter. Finally, a general smooth function of h describes continual change of the parameters.

Let

$$\tilde{W}_{1/2-d_1^0}(\lambda) = \int_0^\lambda s^{-d_1^0} dB(s) \tag{1.11}$$

be a variant of a fractional Brownian Motion with a particular covariance structure,

$$Cov\left(\tilde{W}_{1/2-d_1^0}(\lambda_i), \tilde{W}_{1/2-d_1^0}(\lambda_{i-1})\right) = \frac{\lambda_{i-1}^{1-2d_1^0}}{\Gamma\left(1-d_1^0\right)\left(1-2d_1^0\right)}. \tag{1.12}$$

This differs from the usual fractional Brownian Motion defined as $\int_0^\lambda (\lambda - s)^{-d_1^0} dB(s)$. Further, let

$$B^h(\lambda_i) = B(\lambda_i) - \frac{\pi}{\sqrt{6}} \int_0^{\lambda_i} h_d(u) du \tag{1.13}$$

and

$$\tilde{W}^h_{1/2-d_1^0}(\lambda_i) = \tilde{W}_{1/2-d_1^0}(\lambda_i) - D\left(\lambda_i, d_1^0, h_\mu\right), \tag{1.14}$$

where the second terms reflect the local drift for the break in the memory and in the mean respectively. The latter reads for the general local drift as

$$D\left(\lambda_i, d_1^0, h_\mu\right) = \lim_{T\to\infty} T^{2d_1^0-1} \sum_{t=1}^{[\lambda_i T]} \pi_{t-1}\left(d_1^0 - 1\right) \sum_{j=0}^{t-1} \pi_{j-1}\left(d_1^0\right) h_\mu \left(\frac{t-j}{T}\right).$$

Finally, let

$$F_i^d(\boldsymbol{\lambda},k,1) = \frac{\left(\lambda_i B^h(\lambda_{i+1}) - \lambda_{i+1} B^h(\lambda_i)\right)^2}{\lambda_i \lambda_{i+1}(\lambda_{i+1} - \lambda_i)}, \tag{1.15}$$

$$F_i^\mu(\boldsymbol{\lambda},k,1) = \frac{\left(\lambda_i^{1-2d_1^0} \tilde{W}^h_{1/2-d_1^0}(\lambda_{i+1}) - \lambda_{i+1}^{1-2d_1^0} \tilde{W}^h_{1/2-d_1^0}(\lambda_i)\right)^2}{\lambda_i^{1-2d_1^0} \lambda_{i+1}^{1-2d_1^0}\left(\lambda_{i+1}^{1-2d_1^0} - \lambda_i^{1-2d_1^0}\right)} \text{ and} \tag{1.16}$$

$$F_i^{(d,\mu)}(\boldsymbol{\lambda},k,2) = F_i^d(\boldsymbol{\lambda},k,1) + F_i^\mu(\boldsymbol{\lambda},k,1). \tag{1.17}$$

Theorem 5 provides the asymptotic distribution of the test statistic for breaks in both parameters under the local alternative $H_{1,T}$.

Theorem 5 *(Asymptotic distribution of the test)*
Under Assumptions 1-2 and under $H_{1,T}$,

$$\sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon} F_T^\vartheta(\boldsymbol{\lambda},k;p) \xrightarrow{d} \sup_{\boldsymbol{\lambda}\in\Lambda_\epsilon} \frac{1}{pk} \sum_{i=1}^{k} F_i^\vartheta(\boldsymbol{\lambda},k,p),$$

where the superscript $\vartheta \in \{d, \mu, (d,\mu)\}$ denotes the parameters in which we are testing for breaks.

For the local alternative $H_{1,T}$, the distribution of the test statistic depends on the shape of the $h-$functions and depends therefore on the true break fractions if the h-functions do, e.g. for h being a *stepfunction* in the break fractions λ_i^0.

The asymptotic distribution of the test differs from the one in Bai and Perron (1998) and depends on both standard and fractional Brownian Motion. The terms corresponding to the estimation of memory and mean are additive because of their uncorrelated estimation. If we test for breaks only in the memory, $F_i^d(\boldsymbol{\lambda},k,1)$ corresponds to the one of Bai and Perron (1998) and if we test for breaks only in the mean, the limit distribution $F_i^\mu(\boldsymbol{\lambda},k,1)$ depends on the nuisance parameter d_1^0. $F_i^\mu(\boldsymbol{\lambda},k,1)$ resembles the one for a break in the memory with fractional rather than standard Brownian Motions. In practice, we estimate the memory and compare the test statistic to critical values obtained from simulating the test statistic for a grid of different values of d and fitting a polynomial in d. The validity of this approach follows from Giraitis *et al.* (2003).

Corollary 6 provides the distribution of the test statistic for one break in both parameters under the specific local (one) break hypothesis

$$H_{1,T}' : h_{\vartheta}'(\tau) = \delta_{\vartheta} I\left(\lambda_j^0 \leq \tau\right), \vartheta = \{d, \mu, (d, \mu)\}. \tag{1.18}$$

In this case the local drift of the break in the mean simplifies to

$$D\left(\lambda_i, d_1^0, h_{\mu}'\right) = \delta_{\mu} \frac{\int_0^{\lambda_i} s^{-d_1^0} (s - \lambda_i)^{-d_1^0} ds}{\Gamma\left(1 - d_1^0\right) \sqrt{1 - d_1^0}}. \tag{1.19}$$

Corollary 6 *Under Assumptions 1-2 and under* $H_{1,T}'$,

$$\sup_{\lambda \in \Lambda_{\in}} F_T^{d,\mu}(\boldsymbol{\lambda}, 1; 2) \xrightarrow{d} \sup_{\lambda \in \Lambda_{\epsilon}} \frac{\left[\lambda B(1) - B(\lambda) - \delta_d \frac{\pi}{\sqrt{6}} \left(\lambda\left(1 - \lambda_1^0\right) - \left(\lambda - \lambda_1^0\right)_+\right)\right]^2}{\lambda(1 - \lambda)}$$
$$+ \frac{\left[\lambda \tilde{W}_{1/2 - d_1^0}(1) - \tilde{W}_{1/2 - d_1^0}(\lambda) - \delta_{\mu} \frac{\left(\left(\min\{\lambda, \lambda_1^0\}\right)^{1 - 2d_1^0} \int_{\max\{\lambda, \lambda_1^0\}}^1 s^{-d_1^0} \left(s - \max\left\{\lambda, \lambda_1^0\right\}\right)^{-d_1^0} ds\right)}{\Gamma\left(1 - d_1^0\right) \sqrt{1 - d_1^0}}\right]^2}{\lambda^{1 - 2d_1^0}\left(1 - \lambda^{1 - 2d_1^0}\right)}.$$

The proof follows from substituting $h_{\vartheta}(\tau) = \delta_{\vartheta} I\left(\lambda_j^0 \leq \tau\right), \vartheta = \{d, \mu, (d, \mu)\}$, in Theorem 5. From Corollary 6, because of symmetry, the local power is highest for $\lambda_1^0 = 1/2$. Comparing the local power with the one of the more gradual transition model Chapter 3, we see that for the break in the mean, the local power here is larger.

We focus on tests for *one* break and we simulate the critical values for a grid of d^0 for $\alpha = 0.05$ and $\epsilon = 0.15$. For a break in both parameters they are shown in the first line of Table 1.1 and for a break only in the mean, they are shown in the second line. For a break only in the memory, the critical value corresponds to the one in Bai and Perron (1998), $CV_d = 8.57$.

For establishing the consistency of the test, we have to analyze the estimator using the filter in expression (1.9) under H_1. Similar to Theorems 1 and 2, the break fractions are also consistently estimated at rate T. Thus, we can treat them as if they were known. Next, while the memory estimators $\hat{d}_1, ..., \hat{d}_{k+1}$ are still consistent, the mean estimators $\hat{\mu}_2, ..., \hat{\mu}_{k+1}$ are inconsistent because the applied filters mix observations of the different regimes and converge to weighted averages of the true means of the corresponding and the preceding regimes. Using these results, Theorem 7 provides the consistency of the test for the following alternative

Table 1.1: **Critical Values of F-test for breaks in μ and d and only in μ.**

d^0	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.49
CV	11.6	11.6	11.5	11.5	11.4	11.4	11.4	11.3	11.2	11.2	11.1
CV_μ	8.6	8.5	8.5	8.4	8.4	8.2	8.2	8.1	8.0	7.9	7.9

hypotheses of k breaks at $\boldsymbol{\lambda}^0 = \left(\lambda_1^0, ..., \lambda_k^0\right)$,

$$\begin{array}{r} H_1^d\left(\boldsymbol{\lambda}^0\right) : \theta \neq 0 \text{ and } \nu = 0 \\ H_1^\mu\left(\boldsymbol{\lambda}^0\right) : \theta = 0 \text{ and } \nu \neq 0 \\ H_1^{d,\mu}\left(\boldsymbol{\lambda}^0\right) : \theta \neq 0 \text{ and } \nu \neq 0 \end{array} .$$

Theorem 7 *(Consistency of the test)*

Under Assumptions 1-3 for $k > 1$ breaks,

a) The test for breaks in both parameters diverges at rate T under $H_1^{d,\mu}$ and under H_1^d, and diverges at rate T^{1-2d^0} under H_1^μ.

b) The test for breaks in the memory diverges at rate T under $H_1^{d,\mu}$ and H_1^d.

c) The test for breaks in the mean diverges at rate T^{1-2d^0} under H_1^μ and at rate $T^{1-2\min\{d_i^0\}}$ under $H_1^{d,\mu}$.

Thus, the tests are consistent with a rate of divergence that depends on which parameters are changing and on the memory parameter. In consequence, for a d^0 close to 1/2, the test for a break only in the mean has low power under the alternative.

Finally, if the error term has the stable and known short run dynamics structure ARFIMA(p,d,0) in (1.8), expression (1.13) in Theorem 5 becomes

$$B^h\left(\lambda_i\right) = B\left(\lambda_i\right) - \omega \int_0^{\lambda_i} h_d\left(u\right) du,$$

where $\omega^2 = \frac{\pi^2}{6} - \kappa'\Phi^{-1}\kappa$ is defined in the end of Section 3. A solution to an unknown stable structure is discussed in the empirical application in Section 6.

1.4.2 F-test of ℓ versus $\ell + 1$ breaks

We consider the following hypothesis

$$H_0 : m = \ell \text{ vs. } H_A : m = \ell + 1. \tag{1.20}$$

Technically, we impose ℓ breaks and test each segment for an additional break. The test statistic corresponds to the one in Bai and Perron (1998),

$$F_T\left(\ell+1|\ell\right)=\max_{1\leq i\leq\ell}\frac{1}{\hat{\sigma}_i^2}\left\{S_T(\hat{T}_{i-1},\hat{T}_i)-\inf_{\tau\in\Delta_{i,l}}S_T(\hat{T}_{i-1},\tau,\hat{T}_i)\right\}$$

where

$$\Delta_{i,l}=\left[\tau:\hat{T}_{i-1}+(\hat{T}_i-\hat{T}_{i-1})\eta\leq\tau\leq\hat{T}_i-(\hat{T}_i-\hat{T}_{i-1})\eta\right]$$

and

$$\hat{\sigma}_i^2\xrightarrow{p}\sigma_i^2=\sigma^2.$$

Following the same logic as in the test of *zero* against k breaks, we choose the filter truncated at $\hat{T}_{i-1}$ which is appropriate under H_0 (1.20). The underlying constrained estimator (assuming one regime for the interval $[\hat{T}_{i-1}+1,\hat{T}_i]$) is the one discussed in Theorem 4. For estimating the regime $[\tau,\hat{T}_i]$, the filter is still truncated at $\hat{T}_{i-1}$ rather than at τ and thus differs from the one used in Sections 2-3. Therefore, similar to Section 1.4.1, the mean estimate is not consistent under the alternative. Yet, the test is still consistent.

We consider a local break in regime i. *For* $t=T_{i-1}^0+1,...,T_i^0$,

$$\begin{aligned}H_{1T}^{\ell,i}\quad:\quad d_{i,t}^0&=d_i^0+T^{-1/2}h_d\left(\frac{t-T_{i-1}^0}{T_i^0-T_{i-1}^0}\right)\ \text{and}\\ \mu_{i,t}^0&=\mu_i^0+T^{d_i^0-1/2}h_\mu\left(\frac{t-T_{i-1}^0}{T_i^0-T_{i-1}^0}\right).\end{aligned}$$

There is a local break in regime i with $h_d\left(\tau\right)$ and $h_\mu\left(\tau\right)$ as defined in $H_{1,T}$.

First,

$$T^{-1/2}\sum_{k=1}^{T_{i-1}^0}\left(T^{d_i^0}\sum_{t=1}^{\left[\gamma\left(T_i^0-T_{i-1}^0\right)\right]}\pi_{t-1}\left(d_i^0-1\right)\sum_{l=0}^{t}\pi_l\left(d_i^0\right)\pi_{T_{i-1}^0+t-l-k}\left(-d_i^0\right)\right)u_k \quad (1.21)$$

converges (weakly) to $C\left(\lambda_{i-1}^0,\gamma,d_i^0\right)$, a Gaussian process with mean zero and variance (1.7) with $\lambda_i^0=\gamma T_i^0/T+\left(1-\gamma\right)T_{i-1}^0/T$. Next let $G_{2,\eta}^{h;(d,\mu);(i)}\left(x\right)$ be the distribution function of

$$\sup_{\eta\leq\gamma\leq1-\eta}\left\{\frac{\left(B^h\left(\gamma\right)-\gamma B^h\left(1\right)\right)^2}{\gamma\left(1-\gamma\right)}+\frac{\left(\hat{W}_{1/2-d_i^0,i}^h\left(\gamma\right)-\gamma^{1-2d_i^0}\hat{W}_{1/2-d_i^0,i}^h\left(1\right)\right)^2}{\gamma^{1-2d_i^0}\left(1-\gamma^{1-2d_i^0}\right)}\right\}, \quad (1.22)$$

where $B^h(\gamma)$ is defined in (1.13) and where

$$\hat{W}^h_{1/2-d_i^0,i}(\gamma) = \tilde{W}^h_{1/2-d_i^0}(\gamma) + \frac{\left(1-2d_i^0\right)\Gamma^2\left(1-d_i^0\right)C\left(\lambda_{i-1}^0,\gamma,d_i^0\right)}{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}. \tag{1.23}$$

The first term of (1.23) corresponds to (1.14) with one local break. For $G^{h;d;(i)}_{2,\eta}(x)$, the second term in (1.22) drops and for $G^{h;\mu;(i)}_{2,\eta}(x)$ the first term in (1.22) drops. Theorem 8 provides the asymptotic distribution for testing for a $(\ell+1)$'s break in both parameters.

Theorem 8 *(Asymptotic distribution of the test for ℓ vs. $\ell+1$ breaks)*
Under Assumptions 1,2 and under $H^{\ell,i}_{1T}$,

$$\lim_{T\to\infty} P\left(F_T(\ell+1|\ell)\leq x\right) = \Pi_{j=1}^{\ell+1} G^{h;\vartheta,(j)}_{p,\eta}(x)\,,\ \vartheta\in d,\mu,(d,\mu),$$

where $G^{h=0;\vartheta,(j)}_{p,\eta}(x)\,, j\neq i$.

For the test for a break only in the memory, the test statistic behaves as the one in Bai and Perron (1998). The critical value x_α is the value x for which $G^d_{p,\eta}(x) = \alpha^{1/(l+1)}$ and the critical values are the ones tabulated in Bai and Perron (1998). For the test for a break only in the mean, the distribution depends on the variant of fractional Brownian motion (1.11) plus the additional term coming from applying the too short filter. For this test and for the test for a break in both parameters, $G^{\vartheta,(i)}_{p,\eta}(x)\,,\vartheta=\{\mu,(d,\mu)\}$, differs between the regimes and, consequently, the critical value x_α is the value x for which $\Pi_{i=1}^{\ell+1}G^{\vartheta,(i)}_{p,\eta}(x)=\alpha$. The asymptotic distribution depends on $(d_1^0,...,d_{\ell+1}^0)$ and $(\lambda_1^0,...,\lambda_\ell^0)$. As a consequence, the critical values are obtained on a case-by-case basis given the estimated break partition and memory parameters. Further, the additional term in (1.23) introduces some dependence between the distribution function in the different regimes that has to be taken into account when simulating the critical values. Therefore, using this test is unfeasible in practice. To overcome this problem, we suggest using the bootstrap, which we discuss in the next section.

The consistency and rates of divergence of $F_T(l+1|l)$ follow from using a similar argument as the one for the consistency of the $\sup_\lambda F(\lambda,1,p)$ test for the segment that contains the additional break in Theorem 7.

1.5 Monte Carlo analysis using asymptotic critical values

In this section, we analyze size and power of the three tests discussed in Section 1.4.1, $\sup_{\boldsymbol{\lambda}} F^d_T$, $\sup_{\boldsymbol{\lambda}} F^\mu_T$, $\sup_{\boldsymbol{\lambda}} F^{d,\mu}_T$. For simplicity we analyze the case of one break, using

Table 1.2: **Test for a joint break in memory and mean.**

a) Size. Rejection probabilities when there is no break.

T\d^0	0.05	0.15	0.25	0.35	0.45
200	2.2	6.7	10.0	11.3	13.0
500	3.4	7.5	9.6	9.7	8.8
1000	3.5	6.5	7.0	8.0	7.3

b) Power. Rejection probabilities when there is a break at the half of the sample.

		T=200				T=500			
	$d_2^0\backslash\backslash\mu_2^0$	0.5	**1**	1.5	2	0.5	**1**	1.5	2
d_1^0=0.05	**0.05**	48.2	2.7	50.9	98.1	91.3	3.8	90.9	100.0
	0.10	44.0	4.3	41.7	95.5	83.7	5.4	82.8	100.0
	0.25	45.8	21.8	45.0	84.1	78.7	56.9	81.0	98.7
	0.45	78.6	75.7	80.8	86.5	99.8	99.4	99.2	99.6
d_1^0=0.25	0.05	49.6	21.4	46.8	90.8	85.5	56.5	86.1	99.7
	0.25	22.9	10.7	23.1	54.8	28.8	9.6	27.8	75.3
	0.30	23.1	13.8	23.0	48.9	27.3	13.2	28.0	64.5
	0.45	35.6	31.0	35.7	50.5	65.7	61.5	63.7	73.6
d_1^0=0.45	0.05	83.6	77.6	83.7	91.3	99.8	99.6	99.6	100.0
	0.25	38.9	31.6	38.3	56.8	71.9	61.1	68.4	81.2
	0.40	18.8	16.5	17.9	28.4	17.1	15.1	17.7	27.6
	0.45	16.5	14.8	16.8	26.5	12.3	9.8	12.2	22.4

the critical values provided in Table 1.1. In all following simulations the number of simulations is $1,000$, the distance to the endpoints of the sample $\epsilon = 0.15$, the significance level $\alpha = 0.05$ and the sample sizes are $T = 200, 500$ and $1,000$ for the size and 200 and 500 for the power. We assume an error variance $\sigma^2 = 1$. Since asymptotic results are invariant to the level of the mean, we take $\mu^0 = 1$ if the mean is constant and $\mu_1^0 = 1$ for the mean in the first regime if it is changing. For the size, we analyze $d^0 = 0.05, 0.15, 0.25, 0.35$ and 0.45. For the power, we consider breaks in the memory from $d_1^0 = 0.05$ to $d_2^0 = 0.1$, 0.25 and 0.45, from $d_1^0 = 0.25$ to $d_2^0 = 0.05, 0.3$ and 0.45 and from $d_1^0 = 0.45$ to $d_2^0 = 0.05, 0.25$ and 0.4. Further, we consider breaks in the mean from $\mu_1^0 = 1$ to $\mu_2^0 = 0.5, 1.5$ and 2. The break fraction is always at the half of the sample ($\lambda_1^0 = 0.5$).

First, Table 1.2a) shows the size of a test for a break in both parameters. The estimator of the memory is constrained to lie in the interval $[0, 1/2)$ which naturally has a negative effect on the size in finite samples. This negative effect is largest for $d = 0.05$ and decreases as the sample size increases. For larger memory parameter, the test is oversized in finite samples. This happens because even if the estimation of memory and mean is asymptotically uncorrelated, in finite samples it is still correlated. Table 1.2b) analyzes the power of this test. The power increases in the sample size. In general, a break in the memory is only detectable for larger break sizes. Further, the detectability of a break in the mean decreases considerably in d_0^2

Table 1.3: **Test for a break in the memory.**

a) Size. Rejection probabilities when there is no break.

T\d^0	0.05	0.15	0.25	0.35	0.45
200	1.5	4.2	7.3	9.2	7.4
500	2.3	6.3	8.3	8.4	5.5
1000	2.7	6.5	6.8	7.2	4.8

b) Power. Rejection probabilities when there is a break at the half of the sample.

	d_1^0 =0.05				d_1^0 =0.25				d_1^0 =0.45			
$T\\d_2^0$	**0.05**	0.1	0.25	0.45	0.05	**0.25**	0.3	0.45	0.05	0.25	0.4	**0.45**
200	1.2	1.8	20.8	84.3	25.6	7.7	8.9	29.3	84.5	36.6	10.6	8.7
500	1.2	4.5	64.1	99.6	66.6	8.9	12.3	67.9	99.8	67.9	11.2	7.1

Table 1.4: **Test for a break in the mean.**

a) Size. Rejection probabilities when there is no break.

T\d^0	0.05	0.15	0.25	0.35	0.45
200	6.5	11.2	11.5	11.0	12.8
500	7.3	8.3	8.7	8.2	8.5
1000	7.1	7.1	6.4	7.3	6.0

b) Power. Rejection probabilities when there is a break at the half of the sample.

	T=200				T=500			
$d^0\\\mu_2^0$	0.5	**1**	1.5	2	0.5	**1**	1.5	2
0.05	65.1	7.9	70.8	99.6	95.8	8.5	96.8	100.0
0.25	29.7	12.5	28.2	65.3	36.2	10.0	33.7	82.3
0.45	20.2	14.6	14.9	22.6	15.5	10.0	14.5	22.1

since the higher the true memory in the two regimes, the less precisely the means are estimated. For a non-changing memory of 0.45, the break in the mean is not detected even for larger samples.

Next, we analyze the behavior of the test for a break only in the memory. Table 1.3a) shows the size properties of this test. For $d^0 = 0.05$, the size is too low because of the constrained estimation of the memory. This size distortion vanishes slowly. For larger memory parameters, the test is again slightly oversized. Next, Table 1.3b) shows that the test has power for detecting a break for not too small breaks in the memory. Since the size of a test for a break only in the memory is smaller than the one of a test for a break in both parameters, its power is also smaller.

Finally, we analyze size and power of a test for a break only in the mean. Table 1.4a) displays the size properties of such a test. This test is also slightly oversized. Finally, Table 1.4b) displays the power. Because of the imprecise estimation, a test of a break in the mean has low power when the true memory is close to 0.5. This confirms the lower rate of divergence in Theorem 7.

Table 1.5: **Robustness of tests for a break in one parameter.**

a) Size of test for a break in the memory if there is a break in the mean.

	T=200				T=500			
$d^0 \backslash\backslash \mu_2^0$	0.5	**1**	1.5	2	0.5	**1**	1.5	2
0.05	13.2	6.3	14.0	46.0	20.6	4.5	21.1	79.3
0.25	15.5	8.8	13.2	23.7	10.7	7.9	11.0	22.5
0.45	11.9	11.8	12.3	13.1	6.5	8.7	8.5	8.2

b) Size of test for a break in the mean if there is a break in the memory.

	d_1^0 =0.05				d_1^0 =0.25				d_1^0 =0.45			
$T \backslash\backslash d_2^0$	**0.05**	0.1	0.25	0.45	0.05	**0.25**	0.3	0.45	0.05	0.25	0.4	**0.45**
200	7.9	10.0	19.7	39.2	16.6	13.3	13.6	25.5	23.5	14.3	11.9	14.1
500	8.5	11.7	21.2	41.5	15.2	11.4	12.0	25.6	25.8	12.7	9.1	10.5

1.6 Identifiability of changing parameters

Up to now, we have analyzed the behavior of tests in situations for which they are designed. In this section, we analyze tests for breaks in one parameter for the case that the other parameter is changing. Table 1.5a) shows that the test for a break in the memory is highly oversized if the mean is changing. The reason is that, as mentioned in the end of Section 2, a break in the mean affects the estimation of the memory in finite samples. Table 1.5b) shows that the same is true when testing for a break in the mean if the memory is changing. The reason for this is twofold. First, due to the abrupt change in the memory, the level of the series is changing in the break point of the memory. Second, the mean is estimated at different rates of convergence under the alternative and therefore the difference between SSR_0 and SSR_1^{μ} becomes too large. Therefore, we cannot distinguish between breaks in the memory and breaks in the mean and it is not possible to identify the changing parameter.

First, we focus on testing for a break in the memory when the mean is changing. To solve the mentioned problem we suggest a Chow type test. Let

$$SSR_k'(\boldsymbol{\lambda}) = \min_{\theta_1,\ldots,\theta_{k+1}} \sum_{i=1}^{k+1} \sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \left(\Delta_{t-[\lambda_{i-1}T]}^{d_i} (y_t - \mu_i)\right)^2$$

denote the minimized sum of squares under the alternative of a break in the memory and in the mean given a partition $\boldsymbol{\lambda}$. The estimate of the corresponding break fraction is

$$\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} SSR_k'(\boldsymbol{\lambda}).$$

As in Sections 2 and 3, the filter is truncated at $[\hat{\lambda}_{i-1}T]$ rather than at 1. Next, we use the estimated partition $\hat{\boldsymbol{\lambda}}$ to estimate under the null a constant memory and a changing mean with the corresponding minimized sum of squares

$$SSR_0^d\left(\hat{\boldsymbol{\lambda}}\right) = \min_{d,\mu_1,\dots,\mu_{k+1}} \sum_{i=1}^{k+1} \sum_{t=[\hat{\lambda}_{i-1}T]+1}^{[\hat{\lambda}_i T]} \left(\Delta^d_{t-[\hat{\lambda}_{i-1}T]}\left(y_t - \mu_i\right)\right)^2.$$

For testing for a break in the mean, we estimate under the null of a constant mean and a changing memory with the corresponding minimized sum of squares $SSR_0^{\mu}(\hat{\boldsymbol{\lambda}})$. For simplicity, consider the case of one break. Let

$$F_T^{\vartheta}\left(1;1|\hat{\lambda}_1\right) = \frac{\left(SSR_0^{\vartheta}\left(\hat{\lambda}_1\right) - SSR_k'\left(\hat{\lambda}_1\right)\right)}{SSR_k'\left(\hat{\lambda}_1\right)/(T-2)},\ \vartheta = d,\mu, \tag{1.24}$$

be the test statistic for testing for a break in the memory and the mean respectively.

For testing for a break in the memory ($\vartheta = d$) under the maintained hypothesis of a break in the mean, we assume a local break in the memory and a break in the mean

$$H_{1,T}^{d,\mu_1^0\neq\mu_2^0} : d_t^0 = d_1^0 + T^{-1/2} h_d\left(\frac{t}{T}\right).$$

For testing for a break in the mean ($\vartheta = \mu$) under the maintained hypothesis of a break in the memory, we assume a local break in the mean and a break in the memory

$$\begin{aligned} H_{1,T}^{\mu,d_1^0>d_2^0} &: \ \mu_t^0 = \mu_1^0 + T^{-1/2+d_1^0} h_\mu\left(\frac{t}{T}\right) \text{ or} \\ H_{1,T}^{\mu,d_1^0<d_2^0} &: \ \mu_t^0 = \mu_1^0 + T^{-1/2+d_2^0} h_\mu\left(\frac{t}{T}\right). \end{aligned}$$

Proposition 9a) (b)) discusses the asymptotic distribution of the test for a break in the memory (mean) when the mean (memory) is changing.

Proposition 9 *(Asymptotic distribution of the test for a break in one parameter under the maintained hypothesis of break in other parameter)*

a) Under Assumptions 1-2 and under $H_{1,T}^{d,\mu_1^0\neq\mu_2^0}$,

$$F_T^d\left(1;1|\hat{\lambda}_1\right) \xrightarrow{d} \chi_1^2\left(c_1\right),$$

where $c_1 = \frac{\pi^2}{6}\frac{\left(\lambda_1^0\int_0^1 h_d(u)du - \int_0^{\lambda_1^0} h_d(u)du\right)^2}{\lambda_1^0\left(1-\lambda_1^0\right)}$.

b) Under Assumptions 1-2 and under $H_{1,T}^{\mu,d_1^0>d_2^0}$,

$$F_T^{\mu}\left(1;1|\hat{\lambda}_1\right)\xrightarrow{d}\chi_1^2\left(c_2\right).$$

where $c_2=\frac{D^2\left(\lambda_1^0,d_1^0,h_\mu\right)}{\left(\lambda_1^0\right)^{1-2d_1^0}}$.

c) Under Assumptions 1-2 and under $H_{1,T}^{\mu,d_1^0<d_2^0}$,

$$F_T^{\mu}\left(1;1|\hat{\lambda}_1\right)\xrightarrow{d}\left(1+D_{22}^{\mu}\left(\lambda_1^0,1,d_2^0\right)\right)\chi_1^2\left(c_3\right).$$

where $D_{22}^{\mu}\left(\lambda_1^0,1,d_2^0\right)$ *is defined in (1.6),* $D^2\left(.,d_1^0,h_\mu\right)$ *is defined in (1.19) and* $c_3=\frac{1}{\left(1+D_{22}^{\mu}\left(\lambda_1^0,1,d_2^0\right)\right)}\frac{D^2\left(1,d_1^0,h_\mu\right)}{1-\left(\lambda_1^0\right)^{1-2d_2^0}}$.

First, when there is a break in the mean, the estimator of the break partition $\hat{\boldsymbol{\lambda}}$ converges at rate T to the true break fraction (from Theorem 2). This rate is superconsistent and we can treat the break fraction as known. Therefore, the asymptotic distribution in Parts a) and b) corresponds to the one of a Chow test and the critical values are taken from a χ_1^2.For Part c), because of the too short filter, the asymptotic distribution is not parameter free and we have to simulate the critical values. Finally, if we do not know the direction of the break in the memory, in order to control the size, we choose the critical values from case c) since they are the larger ones. Proposition 9 can be generalized to k breaks. Since the test is also consistent, this procedure makes it possible to distinguish between a break in the memory (mean) and a break in both parameters.

On the other hand, if there are no breaks, $\hat{\boldsymbol{\lambda}}$ converges to a spurious limit and the test statistic asymptotically behaves not as in Proposition 9 but similar to the one in Theorem 5 (the difference comes from the different filters). The critical values from Proposition 9 are not the right ones for this case and we overreject. However, this case only happens with asymptotic probability α (probability of erroneously rejecting $H_0: d_1=d_2$ & $\mu_1=\mu_2$ in the first step). Thus, the size is controlled.

In practice, we can apply the following sequential testing strategy:

1) Test H_0 vs. $H_1: d_1\neq d_2$ and/or $\mu_1\neq\mu_2$ (Corollary 1).
   (i) If do not reject $\rightarrow$ conclude there are no breaks. Stop.
   (ii) If reject $\rightarrow$ conclude there are breaks. $\rightarrow$ 2a) and 2b).
2a) Test $H_0^{d,\mu_1^0\neq\mu_2^0}$ vs. $H_1: d_1\neq d_2$ & $\mu_1\neq\mu_2$ (Prop. 9a))
   (i) If do not reject $\rightarrow$ conclude the memory is not changing.
   (ii) If reject $\rightarrow$ conclude the memory is changing.
2b) Test $H_0^{\mu,d_1^0\neq d_2^0}$ vs. $H_1: \mu_1\neq\mu_2$ & $d_1\neq d_2$ (Prop. 9b)/c))

(i) If do not reject $\rightarrow$ conclude the mean is not changing.
(ii) If reject $\rightarrow$ conclude the mean is changing.

All tests in this sequential procedure are consistent. The size is α for the tests in step 1 and in steps 2a) and 2b) if the respective maintained hypothesis is true. If the mean (memory) is not changing in step 2a) (2b)), the size is $\beta_1 \cdot \alpha$ ($\beta_2 \cdot \alpha$), where β_1 (β_2) denotes the probability of rejecting in the step 2a) (2b)) after having rejected in step 1). This probability lies between α and 1 and depends on the relative strength of the signal in the first step. Therefore, the test of the null of $d_1 = d_2$ *versus* $d_1 \neq d_2$ has size not larger than $\beta_1 \cdot \alpha \leq \alpha$ regardless of the memory and the test of the null of $\mu_1 = \mu_2$ *versus* $\mu_1 \neq \mu_2$ has size not larger than $\beta_2 \cdot \alpha \leq \alpha$.

1.7 Bootstrap

We propose bootstrap procedures for three different situations.

First, we propose the bootstrap as a solution to the encountered size distortions for the test of breaks in mean and/or memory due to constrained estimation for d^0 close to 0 and for a higher memory in Tables 1.2, 1.3 and 1.4. For simplicity, consider the case of *one* break. We apply the following residual bootstrap for testing for breaks in memory and mean:

1. From the estimation under the null, obtain $\hat{d}, \hat{\mu}$ and $\hat{u}_t$.
2. Resample the residuals $\hat{u}_t$ to obtain u_t^*, and generate

$$y_t^* = \hat{\mu} + \Delta_t^{-\hat{d}} u_t^*.$$

3. From the estimation under the null and alternative for the new series y_t^*, obtain the test statistic

$$\sup_{\boldsymbol{\lambda} \in \Lambda_\in} F_T^* (\boldsymbol{\lambda}, k; p) = \sup_{\boldsymbol{\lambda} \in \Lambda_\in} \frac{(SSR_0^* - SSR_k^* (\boldsymbol{\lambda})) / kp}{SSR_k^* (\boldsymbol{\lambda}) / (T - (k+1) p)}, \tag{1.25}$$

with

$$SSR_k^* (\boldsymbol{\lambda}) = \sum_{i=1}^{k+1} \frac{1}{T} \sum_{t=T_{i-1}+1}^{T_i} \left(\Delta_t^{\hat{d}_i^*} (y_t - \hat{\mu}_i^*) \right)^2 .$$

4. Repeat 2-3 B times and obtain from the empirical distribution the bootstrap critical values.

The obtained residuals are asymptotically *close to iid* under H_0. Since the memory is estimated, we integrate the residuals with $\hat{d}$ rather than with d. Therefore, even under H_0 we cannot use a simple resampling under *iid* but we use instead results of Kapetanios (2010), who analyzes the Sieve bootstrap in a similar context, and his remark about the applicability of the CSS estimator. Theorem 10 proves the validity of the bootstrap in the context of no short memory component where the difficulty arises from the fact that the memory is estimated.

Theorem 10 *(Asymptotic behavior of the bootstrap test)*
Under Assumptions 1 and 2 and under H_0 or $H_{1,T}$, the bootstrap based test satisfies

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda}, k, 2\right) \leq x | y_1, ...y_T\right) \xrightarrow{p} P\left(\sup_{\boldsymbol{\lambda}} F\left(\boldsymbol{\lambda}, k, 2\right) \leq x\right)$$

and the test is consistent.

In practice, we use the unconstrained estimator under the alternative rather than the constrained one to obtain the residuals in the first step. By doing so, we expect better power properties. This is valid because of Proposition 11.

Proposition 11 *Under H_0, $0 < \tau < 1/2$,*

$$1) \sup_{\lambda \in [\epsilon, 1-\epsilon]} T^{1/2}\left(d_i\left(\lambda\right) - d^0\right) = O_p\left(1\right), \; i = 1, 2.$$
$$2) \sup_{\lambda \in [\epsilon, 1-\epsilon]} T^{1/2-d^0}\left(\mu_i\left(\lambda\right) - \mu^0\right) = O_p\left(1\right), \; i = 1, 2.$$

Table 1.6a) displays Monte Carlo simulations of the size properties of the bootstrap critical values for testing for a break in both parameters. We apply the Warp bootstrap (Giacomini *et al.*, 2007) for all simulations. Notice that by using the Warp bootstrap, we do not strictly apply the methodology we propose. This can cause some deviations from the nominal size levels. Not surprisingly, the size properties of the test for breaks in both parameters with bootstrap critical values is closer to the nominal level. Table 1.6b) provides the power of this test. For testing for breaks only in the memory and only in the mean, we construct corresponding bootstrap procedures.

Finally, if there are short run dynamics of a stable and known ARFIMA$(p, d, 0)$ structure, the first two steps of the bootstrap change to

1. From the estimation under the null, obtain $\hat{d}, \hat{\mu}, \hat{\alpha}\left(L\right)$ and the residuals

$$\hat{v}_t = \Delta_t^{\hat{d}} \hat{\alpha}\left(L\right)\left(y_t - \hat{\mu}\right).$$

Table 1.6: **Bootstrap test for a break in memory and mean.**

a) Size. Rejection probabilities when there is no break.

T\\d^0	0.05	0.15	0.25	0.35	0.45
200	6.4	6.8	6.0	5.1	5.2
500	4.2	5.8	4.8	4.6	6.3
1000	5.3	5.6	5.3	4.0	4.9

b) Power. Rejection probabilities when there is a break at the half of the sample.

		T=200				T=500			
	$d_2^0 \diagdown \mu_2^0$	0.5	**1**	1.5	2	0.5	**1**	1.5	2
d_1^0=0.05	**0.05**	50.1	6.4	53.5	97.6	90.8	5.8	89.8	100.0
	0.10	42.2	5.7	44.6	91.8	79.6	5.5	82.7	99.9
	0.25	36.1	21.8	32.7	74.1	74.2	45.9	77.5	97.3
	0.45	63.7	62.5	71.0	77.8	98.9	99.1	98.6	99.1
d_1^0=0.25	0.05	41.6	20.6	40.7	86.2	78.6	48.7	84.7	98.8
	0.25	17.4	6.4	15.2	42.9	16.1	6.5	23.3	57.7
	0.30	14.9	9.5	17.1	41.5	15.0	7.7	19.4	47.6
	0.45	22.0	20.5	23.2	32.5	52.6	50.1	54.7	61.0
d_1^0=0.45	0.05	72.9	67.3	71.2	85.7	99.5	99.0	99.7	99.8
	0.25	26.6	18.3	25.1	41.0	56.2	49.8	60.9	62.5
	0.40	10.9	9.2	10.2	15.9	8.9	10.3	12.7	15.6
	0.45	7.7	6.6	8.6	17.1	6.4	5.6	10.7	11.7

2. Resample the residuals $\hat{v}_t$ to obtain v_t^* and generate

$$y_t^* = \hat{\mu} + \hat{\alpha}^{-1}(L)\,\Delta_t^{-\hat{d}} v_t^*.$$

Second, we analyze a bootstrap procedure for a test for a break in the memory (mean) that is robust to the presence of a break in the mean (memory). Such a test is necessary since the tests defined in Theorem 4 suffer from the size distortions shown in Table 1.5, and the tests in Proposition 9 require a break in the not tested parameter. For the test for a break in the memory that is robust to the presence of a break in the mean, we apply the following residual bootstrap:

1. From the estimation under the null, minimizing SSR_0^{μ}, obtain $\hat{\lambda}_1, \hat{d}, \hat{\mu}_1, \hat{\mu}_2$ and the residuals $\hat{u}_t$. In line with the procedure described in Sections 2-3, use the filter (1.4).

2. Resample the residuals $\hat{u}_t$ to obtain u_t^*, and generate

$$y_t^* = \begin{cases} \hat{\mu}_1 + \Delta_t^{-\hat{d}} u_t^*, & t \leq \hat{\lambda}_1 T \\ \hat{\mu}_2 + \Delta_t^{-\hat{d}} u_t^*, & t > \hat{\lambda}_1 T. \end{cases}$$

3. From the estimation under the null and the alternative for y_t^*, obtain a bootstrap version of the test statistic (1.24).

4. Repeat 2-3 B times and obtain from the empirical distribution the bootstrap critical values.

Proposition 12 discusses validity and consistency of the bootstrap procedures in both cases. If the not tested parameter is not changing, the behavior follows from combining Theorem 10 and Proposition 11. If the not tested parameter is changing, the behavior follows from similar arguments as the ones in Proposition 9.

Proposition 12 *(Asymptotic behavior of the robust bootstrap test)*

a) Under Assumptions 1-2, for testing for a break in the memory, the bootstrap based test, corresponding to (1.25), satisfies under H_0 *and* $H_{1,T}$*,*

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda},k,1\right) \leq x|y_1,...,y_T\right) \xrightarrow{p} P(\sup_{\boldsymbol{\lambda}} F^d\left(\boldsymbol{\lambda},k,1\right) \leq x),$$

under $H_{1,T}^{d,\mu_1^0\neq\mu_2^0}$,

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda},k,1\right) \leq x|y_1,...,y_T\right) \xrightarrow{p} \chi_1^2.$$

Further, the test is consistent.

b) Under Assumptions 1-2, for testing for a break in the mean, the bootstrap based test satisfies under H_0 *and* $H_{1,T}$*,*

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda},k,1\right) \leq x|y_1,...,y_T\right) \xrightarrow{p} P\left(\sup_{\boldsymbol{\lambda}} F^\mu\left(\boldsymbol{\lambda},k,1\right) \leq x\right).$$

and under $H_{1,T}^{\mu,d_1^0>d_2^0}$,

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda},k,1\right) \leq x|y_1,...,y_T\right) \xrightarrow{p} \chi_1^2$$

and under $H_{1,T}^{\mu,d_1^0<d_2^0}$,

$$P\left(\sup_{\boldsymbol{\lambda}} F_T^*\left(\boldsymbol{\lambda},k,1\right) \leq x|y_1,...,y_T\right) \xrightarrow{p} \left(1+D_{22}^{\mu}\left(\lambda_1^0,1,d_2^0\right)\right)\chi_1^2$$

Further, the test is consistent.

$F^d\left(\boldsymbol{\lambda},k,1\right)$ *and* $F^\mu\left(\boldsymbol{\lambda},k,1\right)$ *are both defined in Theorem 5.*

As discussed in Section 1.4.1, the asymptotic distribution of the test statistic for testing for a break in the memory (mean) differs between the case when the mean (memory) changes and the case when it does not change. The bootstrap based test has to take this into account and converges in probability to the corresponding asymptotic distributions. If there is a break in the mean, $\hat{\lambda}_1$ converges to the true break fraction and due to the superconsistency, the test behaves as a Chow test (Proposition 9). If there is no break in the mean, $\hat{\lambda}_1$ has a spurious limit and the asymptotic behavior corresponds to the first term of Corollary 1. Table 1.7a) displays the size of this alternative bootstrap procedure.It turns out that the test is

Table 1.7: **Size of robust bootstrap tests.**

a) Size of a bootstrap test for a break in d that is robust to a break in μ.

	T=200				T=500				T=1000			
$d^0 \backslash\backslash \mu_2^0$	0.5	**1**	1.5	2	0.5	**1**	1.5	2	0.5	**1**	1.5	2
0.05	6.9	7.8	6.5	7.9	7.9	7.2	4.6	5.4	7.0	7.3	6.5	5.7
0.15	7.2	10.0	7.4	7.4	6.9	6.9	6.9	6.8	4.7	5.9	6.0	5.9
0.25	8.9	7.8	6.3	6.2	6.2	8.0	6.6	4.9	6.3	6.2	6.1	4.9
0.35	7.9	6.9	6.9	5.9	4.8	7.7	4.8	4.9	7.0	6.7	7.0	3.4
0.45	4.9	6.9	4.5	5.2	4.3	6.1	4.5	4.3	5.5	5.5	5.9	3.7

b) Size of a bootstrap test for a break in μ that is robust to a break in d.

	d_1^0 =0.05				d_1^0 =0.25				d_1^0 =0.45			
$T \backslash\backslash d_2^0$	**0.05**	0.1	0.25	0.45	0.05	**0.25**	0.3	0.45	0.05	0.25	0.4	**0.45**
200	4.1	9.4	4.2	9.7	5.8	6.5	4.0	10.8	6.0	5.0	4.4	7.0
500	6.8	6.4	6.0	10.1	6.9	7.1	5.0	8.2	5.7	4.5	4.8	8.7
1000	6.2	5.2	6.7	10.2	4.0	4.8	8.4	10.3	4.0	5.5	8.2	6.3

still slightly oversized when the mean is not changing (2nd column). In this case, we estimate a changing mean with a spurious break point. Thus, the generated series has a changing mean at this spurious break point and we frequently estimate a break at this point. For larger sample sizes, the size gets closer to the nominal level. The power is clearly larger than the one for an alternative conservative strategy of using always critical values from Theorem 5. This robust bootstrap test also improves steps 2a) and 2b) in the sequential procedure in Section 1.6. Table 1.7b) provides the size of the test for a break in the mean that is robust to the break in the memory. The test is still oversized when the memory is close to 0.5 since in this case, the mean is imprecisely estimated.

Finally, we present a bootstrap procedure for testing ℓ versus $\ell + 1$ breaks to solve the problems described in the previous section. For simplicity, consider the case of *one* vs. *two* breaks. We apply the following residual bootstrap:

1. From the estimation under the null H_{1T}^{ℓ}, which corresponds to the methodology described in Sections 2-3, obtain $\hat{\lambda}_1, \hat{d}_1, \hat{d}_2, \hat{\mu}_1, \hat{\mu}_2$ and the residuals $\hat{u}_t$.

2. Resample the residuals $\hat{u}_t$ to obtain u_t^*, and generate

$$y_t^* = \hat{\mu}_i + \Delta_t^{-\hat{d}_i} u_t^*, \text{ for } \hat{\lambda}_{i-1}T + 1, ..., \hat{\lambda}_i T.$$

3. From the estimation under the null and under the alternative for the new series y_t^*, obtain the test statistic from Theorem 5.

4. Repeat 2-3 B times and obtain from the empirical distribution the bootstrap critical values.

This bootstrap test is valid for similar reasons as the ones in Theorem 10 and avoids the problem of obtaining the asymptotic critical values on a case by case basis.

1.8 Empirical Application

In the previous sections, we have assumed that the short run dynamics structure is known. For the empirical application this assumption has to be relaxed. Since the consistency of the parametric memory estimation depends on the knowledge of this autoregressive structure, we need a preliminary estimate of the memory. For a stable fractionally integrated progress, Hualde and Robinson (2011) suggest using the following approach: First, obtain a preliminary memory estimate from a semiparametric estimation (e.g. the local Whittle estimator (Robinson, 1995)) and use this estimate to filter the series to obtain (approximately) short memory. Next, choose the orders p, q of the short memory $ARMA\left(p,q\right)$ structure by minimizing an information criterion. Finally, the parameters of the $ARFIMA\left(p,d,q\right)$ are estimated parametrically.

In our case, we need to obtain the preliminary semiparametric estimate under the alternative rather than under the null. Thus, as in Hsu (2005) and Hassler and Meller (2011), we use a modified version of the Exact Local Whittle estimator (Shimotsu and Phillips, 2005, Shimotsu, 2010) and we further modify it by allowing also for a break in the memory. In particular, we define the periodogram and the discrete Fourier transform of a time series $\{x_t\}_{t=t_1}^{t_2}$ evaluated at the fundamental frequencies as

$$I_x\left(v_j\right) = \left|\zeta_x\left(v_j\right)\right|^2$$

and

$$\zeta_x\left(v_j\right) = \left(2\pi T\right)^{-1/2}\sum_{t=t_1}^{t_2} x_t e^{itv_j}, v_j = \frac{2\pi j}{t_2 - t_1}.$$

Given a break fraction λ, the mean estimators are

$$\mu_1\left(\lambda\right) = \frac{1}{[\lambda T]}\sum_{t=1}^{[\lambda T]} y_t \text{ and } \mu_2\left(\lambda\right) = \frac{1}{[(1-\lambda)\,T]}\sum_{t=[\lambda T]+1}^{T} y_t.$$

The memory estimator is

$$\hat{d}_i\left(\lambda\right) = \arg\min_{d_i} R\left(d_i, \lambda\right),$$

where for $n_T^{(1)} = (\lambda T)^{\alpha}$ and $n_T^{(2)} = ((1-\lambda)T)^{\alpha}$ $0 < \alpha < 1$,

$$R(d_i, \lambda) = \log \hat{G}(d_i, \lambda) - 2d_i \frac{1}{n_T^{(i)}} \sum_{j=1}^{n_T^{(i)}} \log v_j$$

and

$$\hat{G}(d_i, \lambda) = \frac{1}{n_T^{(i)}} \sum_{j=1}^{n_T^{(i)}} I_{u^{(i)}(\lambda)}(v_j),$$

where

$$\begin{array}{l} u_t^{(1)}(\lambda) = \Delta_t^{d_1}(y_t - \mu_1(\lambda)),\ t \leq [\lambda T] \\ u_t^{(2)}(\lambda) = \Delta_{t-[\lambda T]}^{d_2}(y_t - \mu_2(\lambda)),\ t > [\lambda T] \end{array}. \tag{1.26}$$

Finally, the break fraction is estimated as

$$\hat{\lambda} = \arg\min_{\lambda} \{R(d_1(\lambda), \lambda) + R(d_2(\lambda), \lambda)\}.$$

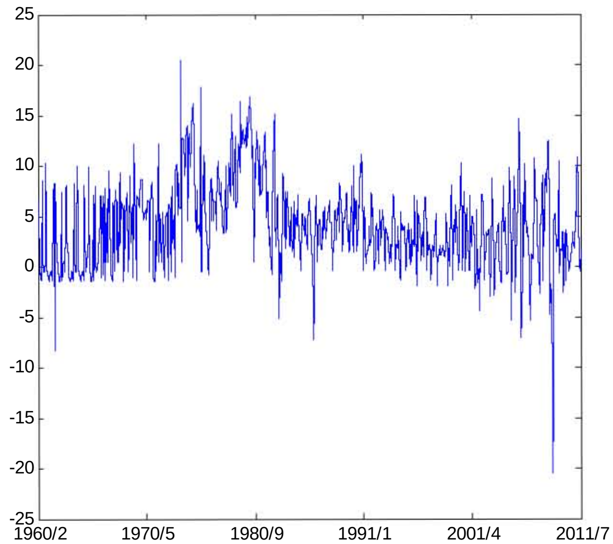
From Lavielle and Ludeña (2000), such a break fraction estimator should estimate the break fraction at rate n_T. The subsequent estimators of the parameters in the two regimes behave as described in Shimotsu (2006). In the following, we choose $\alpha = 0.7$.

We filter the data using the semiparametric estimates $(\tilde{d}_1, \tilde{d}_2, \tilde{\mu}_1, \tilde{\mu}_2, \tilde{\lambda}_1)$ to obtain residuals that are close to $I(0)$. Then, we determine p in the $AR(p)$ structure using the Bayesian information criterion (BIC). Afterwards, we employ the parametric testing procedure described in Section 4 and 1.6. The extension to more breaks is straightforward.

If the short run dynamics is also changing, yet with a stable structure, we include $\alpha_1(L)$ and $\alpha_2(L)$ in the parametric estimation. This adds another dimension to the test, along the lines of Boldea and Hall (2010). The first component (1.13) consists for a changing AR(p) of a $1+p$ dimensional Brownian Motion. Because the pre-estimation is semiparametric, we need to assume that $\alpha(L)$ is changing at the same point as the memory and/or the mean. In the following, we assume that $\alpha(L)$ and the memory are changing at the same time.

Next, we illustrate how the procedure works for a real data set. We consider the U.S. inflation time series which is already extensively analyzed in the literature. Kumar and Okimoto (2007) discuss that inflation persistence can be measured in different ways, I(0) vs I(1), largest autoregressive root, sum of autoregressive coefficients and I(d). They argue that the latter is preferable because of its flexibility. The literature is inconclusive about whether inflation is stationary, fractionally in-

Figure 1-1: **Seasonally Adjusted Monthly US Inflation**



tegrated or has a unit root and whether or not it has breaks in the deterministic part and/or the memory (See Martins and Rodrigues (2010) and Hassler and Meller (2011) for a good summary of the empirical literature). Hsu (2005) finds two breaks in the mean in January 1973 and September 1981 when allowing for fractionally integrated errors. Hassler and Scheithauer (2011) and also Sibbertsen and Kruse (2009) find a break from a unit root to a memory smaller than *unity* in the first quarter of 1982. Hassler and Meller (2011) conclude that there is one (or possibly two) break(s) in the memory. Mayoral (2012) concludes that the U.S. inflation is a fractionally integrated series with a memory around 0.6, though without testing for breaks in the memory parameter. Martins and Rodrigues (2010) find a break from a unit root to around 0.3 in July 1982, yet without taking into account potential breaks in the mean.

As in Hassler and Meller (2011), we analyze the monthly U.S. CPI data collected by the Organization for Economic Cooperation and Development (OECD). This series comprises 619 observations from January 1960 until July 2011. Inflation is computed as

$$\pi_t = 1200 \log \left(CPI_t / CPI_{t-1} \right).$$

Finally, we seasonally adjust the series by subtracting seasonal means and adding the overall mean. Figure 1-1 displays the seasonally adjusted inflation series.

First, we apply the semiparametric procedure and find two breaks in November 1972 and in August 1981. Table 1.8a) displays memory and mean estimates in the regimes and the Bayesian information criterion (BIC) of AR(p) models for the filtered data in the regimes. Thus, we choose a AR(1) structure for the filtered data. Next, we apply the parametric testing procedure with an underlying ARFIMA(1,d,0) structure. In a first step, we determine sequentially the number of breaks in the memory and autoregressive parameter and/or the mean allowing for fractionally integrated errors under H_0 and H_1. In a second step, we identify whether the breaks are in the memory and/or the mean. Because of the size distortions mentioned in Section 1.5, we compare the test statistic to the bootstrap critical values. It turns out that for this data, the bootstrap critical values differ considerably from the asymptotic ones. The candidate for the first break is October 1981 and we reject the hypothesis H_0 of no break at the 1% significance level. Thus, there is at least *one* break in October 1981. In the same way, we next test, whether there is an additional break in the periods before and after October 1981. Notice that we choose the parameter ϵ sample size dependent, leaving always at least 50 observations on both sides of the sample. Table 1.8b) displays the sequential tests for the number of breaks, the estimated break points, the test statistics and the bootstrap critical values. We conclude that there are two breaks, one in February 1973 and one in October 1981. The former, corresponds to the first oil crisis and the latter corresponds to the Volcker disinflation period, the end of the second oil crisis and the great moderation. The potential break in September 1990 is not found to be significant. Table 1.8c) summarizes the estimates of memory (with standard errors), mean and autoregressive parameter for the three regimes. At the first oil shock, the persistence increases, and along with the Volcker disinflation and great moderation the persistence decreases considerably. This is in line with arguments that the U.S. adopted an implicit inflation targeting (see Goodfriend, 2004). Thus, a decrease in the persistence contributes to the stabilization of the U.S. economy in this period. Further, in this analysis, we do not find a Greenspan effect in the 90s.

In the second step, we use the methodology in Proposition 9 to determine which parameters are the changing ones for each break point. Table 1.8d) provides test statistics and bootstrap critical values for testing for a break in the memory (mean) under the maintained hypothesis of a break in the mean (memory). We conclude that both breaks are in the mean but only the one in October 1981 is also in the memory. Therefore, we reestimate a constant memory and autoregressive parameter for the period 1960:01 to 1981:10 ($\hat{d} = 0.30$ (0.07) and $\hat{\alpha} = -0.29$).

Our memory estimates are considerably lower than the estimates in Martins and Rodrigues (2010), Hassler and Scheithauer (2011), Sibbertsen and Kruse (2009) and Mayoral (2012). However, these papers do not allow for breaks in the mean and, therefore, their memory estimates might be spuriously high. Hassler and Meller (2011) allow for breaks in the memory and obtain similar memory estimates as ours. However, they test for breaks in mean and memory sequentially rather than

Table 1.8: **Breaks in US Inflation Rate**

a) Semiparametric pre-estimation: Memory, mean and BIC for order of AR(p)

Period	d	μ	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)
1960:02-1972:12	0.19	2.91	2.74	**2.69**	2.72	2.73	2.76	2.78
1973:01-1981:08	0.48	8.90	2.91	**2.76**	2.79	2.82	2.86	2.90
1981:09-2011:07	0.12	3.00	2.58	**2.53**	2.53	2.55	2.57	2.59

b) Sequential procedure: F-tests for breaks in the three parameters.

Test	Break point	F	$CV^*_{0.95}$ $(CV^*_{0.99})$
0 vs 1	1981:10	**55.50**	35.83 (41.60)
1 vs 2	1973:02	**25.09**	18.33 (22.84)
2 vs 3	1990:09	13.64	16.30

c) Parameter estimates in the regimes.

Period	d	μ	α
1960:02-1973:02	0.27 (0.09)	3.08	-0.31
1973:03-1981:10	0.42 (0.11)	9.74	-0.25
1981:11-2011:07	-0.07 (0.07)	2.98	0.44

d) Sequential procedure: F-tests for identifying the changing parameter.

	Break in d and α		Break in μ	
Break point	F	$CV^*_{0.95}$	F	$CV^*_{0.95}$
1973:02	4.70	6.65	**9.82**	6.35
1981:10	**16.85**	6.58	**13.06**	5.89

simultaneously. By testing for breaks in mean and memory simultaneously, we reduce spurious effects caused by the finite sample correlation between the respective estimates.

1.9 Final Remarks

The analysis is extendable in several directions. First, we have analyzed breaks in (asymptotically) stationary time series with $0 \leq d_j^0 < 1/2$. The analysis can be extended to a memory in the interval $-1/2 < d_j^0 \leq 0$. In this case, the stronger signals come from the break in the mean rather than the break in the memory. Nevertheless, this is still too restrictive for many applications. For example, assume a series with a linear trend and with a nonstationary memory with $1/2 < d_j^0 \leq 1$ or $1 \leq d_j^0 < 3/2$,

$$y_t = \mu_j^0 + \beta_j^0 t + \Delta_t^{-d_j^0} u_t, \ t = T_{j-1}^0 + 1, ..., T_j^0.$$

In this case, we apply a first-differencing filter to the process to obtain

$$\Delta y_t = \beta_j^0 + \Delta_t^{1-d_j^0} u_t, \ t = T_{j-1}^0 + 1, ..., T_j^0.$$

The differenced process has a a changing mean and a new changing stationary memory parameter, $d_j^0 - 1 \in (-1/2, 0)$ for $1/2 < d_j^0 \leq 1$ and $d_j^0 - 1 \in (0, 1/2)$ for $1 \leq d_j^0 < 3/2$, for which our methodology is valid. Note that the original mean cannot be estimated and breaks in it are not identifiable and do not contribute to finding the break. Taylor *et al.* (2010) propose a test for a break in the mean that is robust for any d, including nonstationary ones. Next, if the process has a changing mean and linear trend and a memory lying in Θ, the analysis increases by one further dimension. This analysis is beyond the scope of this paper.

In the previous analysis, we have assumed that the error follows (2.1). However, this so called Type II long memory process is not the only possibility of defining a long memory process. Alternatively, we could assume a Type I long memory error

$$\Delta_{\infty}^{-d_j^0} u_t = \sum_{j=0}^{\infty} \pi_j \left(d_j^0\right) u_{t-j}, 0 \leq d_j^0 < 1/2.$$

The estimation of the memory and of the short run dynamics is unaffected. The mean estimation, on the other hand, has an additional term that is similar to (1.6). In the tests, the variance is increased in a similar way as in Theorem 8. This increased variance would have to be taken into account. Further, since the mean is less precisely estimated, the resulting local power would be slightly lower.

In the previous analysis, we have assumed a not breaking variance. To relax this assumption, we could robustify the testing procedure, by applying a Wald test with a heteroskedastic robust estimate of the covariance matrix (see Hassler and Meller, 2011). Alternatively, we can incorporate breaks in the variance into the procedure. We would then test for breaks in the mean, the memory and the variance. Zhou and Perron (2008) derive how to test for non simultaneous breaks in mean and in variance.

Finally, we have assumed one of two situations. Breaks are exclusively in one parameter or always simultaneously in both parameters. Nevertheless, the proposed procedure also works if the breaks are not simultaneous. Assume the true process has k_1 breaks in the memory and k_2 breaks in the mean at potentially different break points. Using the sequential testing in the lines of Bai and Perron (1998), we first detect $k = k_1 + k_2$ breaks. Next, using the sequential procedure in Section 1.6, we obtain for each of the k breaks, whether it is a break in the memory, in the mean or in both parameters.

1.10 Appendix A: Lemmata and Propositions

Lemmata

Lemma 13 *Denote for*$[\lambda_{i-1}T] < t \leq [\lambda_i T]$,

$$d_t\left(\lambda_{i-1},\theta_i\right) = \hat{u}_t\left(\lambda_{i-1},\theta_i\right) - u_t\left(\lambda_{i-1},\theta_i^0\right). \tag{1.27}$$

Under Assumptions 1-3, uniformly in $\theta \times r \in \Theta \times [0,1]$ *and in* s *for* $s - \lambda_{i-1}^0 = O\left(T^{-1}\right)$,

$$a)\ T^{-\delta_i}\sum_{[sT]+1}^{[rT]} d_t^2\left(s,\theta\right) = O_p(1)$$

$$b)\ T^{-\delta_i}\sum_{[sT]+1}^{[rT]} u_t\left(s,\theta^0\right) d_t\left(s,\theta\right) = o_p(1)$$

Proof. We have to show uniform convergence of $\sum_{t=[sT]+1}^{[rT]} d_t^2\left(s,\theta\right)$ and $\sum_{t=[sT]+1}^{[rT]} u_t\left(s,\theta^0\right) d_t\left(s,\theta\right)$ for $\left(s-\lambda_{i-1}^0\right) = O\left(T^{-1}\right)$. The proofs of tightness use among other Lemma 15 and 16 of Johansen and Nielsen (2010). For Part a), we provide a sketch of the proof in the supplemental Appendix. ■

Lemma 14 *If* $\lambda_i^{(1)} < \lambda_i^0$, *for some* i *then*

$$\begin{aligned}
(i)\sup{}_{\lambda_i^{(1)}<\lambda_i^0} T^{-\delta_i}\sum_{t=1}^{T} u_t\left(s,\theta^0\right) d_t\left(s,\theta\right) &= o_p\left(1\right)\\
(ii)\liminf P\left[T^{-\delta_i}\sum_{t=1}^{T} d_t^2\left(s,\theta\right) > C\right] &> \epsilon,\ \textit{for some}\ C>0,\ \epsilon>0.
\end{aligned}$$

For a break at T_i^0 *in the memory and the mean or only in the memory* : $\delta_i = 1$ *and for a break only in the mean:* $\delta_i = 1 - 2d_i^0$.

Proof. We have to show that for any break fraction smaller than the true one, $\lambda_i^{(1)} < \lambda_i^0$, the term $T^{-\delta_i}\sum_{t=1}^{T} u_t d_t$ vanishes and $T^{-\delta_i}\sum_{t=1}^{T} d_t^2$ is of order $O_p^+\left(1\right)$.

ii) Assume m breaks and consider the break in $\lambda_i^0 T$ in (d,μ) or d. For $\lambda_i^{(1)} < \lambda_i^0$, we know from Lemma 13 that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} d_t^2\left(\lambda_{i-1},\theta_i\right) &\geq \frac{1}{T}\sum_{t=\left[\lambda_{i-1}^{(1)}T\right]+1}^{\left[\lambda_i^0 T\right]} d_t^2\left(\lambda_{i-1},\theta_i\right) + \frac{1}{T}\sum_{t=\left[\lambda_i^0 T\right]+1}^{\left[\lambda_i^{(1)}T\right]+1} d_t^2(\lambda_{i-1},\theta_i)\\
&\xrightarrow{p} \left(\lambda_i^0 - \lambda_{i-1}^0\right)\sigma^2\sum_{j=1}^{\infty}\pi_j^2(d_i - d_i^0) + \left(\lambda_i^{(1)} - \lambda_i^0\right)\sigma^2\sum_{j=1}^{\infty}\pi_j^2(d_i - d_{i+1}^0)
\end{aligned}$$

Similarly as in Boldea and Hall (2010), we can choose an η small enough so that the previous term bounds

$$\begin{aligned} & \eta\sigma^2 \inf_{d_i}\left[\sum_{j=1}^{\infty}\pi_j^2(d_i-d_i^0)+\sum_{j=1}^{\infty}\pi_j^2(d_i-d_{i+1}^0)\right] \\ \geq \; & \eta\sigma^2\left[(d_i-d_i^0)^2\sum_{j=1}^{\infty}\pi_j^2(0)+(d_i-d_{i+1}^0)^2\sum_{j=1}^{\infty}\pi_j^2(0)\right] \\ > \; & \eta\sigma^2\left(\frac{\pi^2}{6}-1\right)\left[(d_i-d_i^0)^2+(d_i-d_{i+1}^0)^2\right]>0 \end{aligned}$$

uniformly in d_i.

Next, we consider the consistency of the break fraction estimator, when there is only a break in the mean. For $d_i^0>0$, d_i and d_{i+1} converge at rate $T^{1/2}$ to d_i^0 and terms including $(d_j-d_i^0)$ vanish. From the proof of Lemma 13,

$$\begin{aligned} T^{2d_i-1}\sum_{t=1}^{T}d_t^2 &\geq T^{2d_i-1}\sum_{t=\left[\lambda_{i-1}^{(1)}T\right]+1}^{\left[\lambda_i^0T\right]}d_t^2\left(\lambda_{i-1}^{(1)},\theta_i\right)+T^{2d_i-1}\sum_{t=\left[\lambda_i^0T\right]+1}^{\left[\lambda_i^{(1)}T\right]}d_t^2\left(\lambda_{i-1}^{(1)},\theta_i\right) \\ &\geq T^{2d_i-1}\sum_{t=\left[\lambda_{i-1}^{(1)}T\right]+1}^{T_i^0}\left[\left(\mu_i^0-\mu_i\right)\Delta_{t-\lambda_{i-1}^{(1)}T}^{d_i}1\right]^2 \\ &+T^{2d_i-1}\sum_{t=T_i^0+1}^{\left[\lambda_i^{(1)}T\right]}\left[\left(\mu_{i+1}^0-\mu_i\right)\Delta_{t-T_i^0}^{d_i}1+\left(\mu_i^0-\mu_i\right)\left(\Delta_{t-\lambda_i^{(1)}T}^{d_i}1-\Delta_{t-T_i^0}^{d_{i+1}}1\right)\right]^2. \end{aligned}$$

First, both terms have a nonnegative limit. The first term's limit equals zero only if $(\mu_i^0-\mu_i)=o_p(1)$. But in this case, the second term's limit is larger than zero. Therefore, uniformly in μ_i and d_i for $(d_i-d_i^0)=O_p(T^{-1/2})$, the term is positive.

Note that for the contradiction established for the break in T_i^0, the less favorable case is the one where all other breaks $j\neq i$ are consistently estimated at the rate established in Theorem 2. Therefore, it suffices to consider this case.

i) Follows from Lemma 13. ■

Lemma 15 states some properties for the regressor function and its derivative that are needed in the proofs. In Boldea and Hall (2010), they are assumed in their Assumptions 2-4. In our context, they are a consequence of Assumption 1 and 2.

Lemma 15 *Recall $f_t(\theta)=(\Delta_{t-T_{i-1}}^{d_i}-1)(y_t-\mu_i)$ and define $F_t(\theta)=\frac{\partial f_t(\theta)}{\partial\theta}$, a px1 vector, a function of θ_i for $t\in[T_{i-1}+1,T_i]$ and $F_{k,t}(\theta)$, $k=d,\mu$ the derivative with respect to d and μ respectively. Further, define $\bar{T}(d_i^0)=(diagT^{-1/2},T^{d_i^0-1/2})$*

a) *Given the superconsistent rate of convergence of the break fractions, $S_{i,T}(\lambda_{i-1},\lambda_i,\theta_i)$ defined in (1.5), appropriately standardized, converges to a limit that is minimized in $d_i=d_i^0$ and $\mu_i=\mu_i^0$.*

b) *Evaluated at the true θ_i^0 and the true break fractions,*

$$\begin{aligned} D_{T,i}\left(\theta_i^0\right) &= \bar{T}\left(d_i^0\right)\sum_{t=T_{i-1}^0+1}^{T_i^0} F_t\left(\lambda_{i-1}^0,\theta_i^0\right)F_t\left(\lambda_{i-1}^0,\theta_i^0\right)'\bar{T}\left(d_i^0\right) \\ &\xrightarrow{p} \sigma^2 D_i^0\left(\lambda_{i-1}^0,\lambda_i^0,\theta_i^0\right) \end{aligned}$$

where

$$D_i^0\left(\lambda_{i-1}^0,\lambda_i^0,\theta_i^0\right)=\begin{pmatrix} \frac{\pi^2}{6}\left(\lambda_i^0-\lambda_{i-1}^0\right) & 0 \\ 0 & \frac{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}{\left(1-2d_i^0\right)\Gamma^2\left(1-d_i^0\right)} \end{pmatrix}.$$

c) *Uniformly in (s,r,θ) for $\left(s-\lambda_{i-1}^0\right)=O_p\left(T^{-1}\right)$ and $r>\lambda_i^0$,*

$$D_{i,T}\left(\theta_i\right)=\bar{T}\left(d_i\right)\sum_{t=[sT]+1}^{[rT]} F_t\left(s,\theta\right)F_t\left(s,\theta\right)'\bar{T}\left(d_i\right)\xrightarrow{p}\sigma^2 D_i\left(s,r,\theta\right)$$

where

$$D_i\left(r,\theta\right)=\begin{pmatrix} \left(r-\lambda_i^0\right)\sigma^2\sum\limits_{j=0}^{\infty}\dot{\pi}_j^2(d-d_{i+1}^0)+\left(\lambda_i^0-\lambda_{i-1}^0\right)\sigma^2\sum\limits_{j=0}^{\infty}\dot{\pi}_j^2(d-d_i^0) & 0 \\ 0 & \frac{\left(r-\lambda_{i-1}^0\right)^{1-2d}}{(1-2d)\Gamma^2(1-d)} \end{pmatrix}.$$

d) *Evaluated at the true d_i^0 and the true break fractions*

$$A_i\left(\theta_i^0\right)=Var\left[diag\left(T^{-1/2},T^{d_i^0-1/2}\right)\sum\nolimits_{t\in I_i^0}u_t\left(\lambda_{i-1}^0,\theta_i^0\right)F_t\left(\lambda_{i-1}^0,\theta_i^0\right)\right]\xrightarrow{p}A\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)$$

where

$$A\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)=\begin{pmatrix} \sigma^4\frac{\pi^2}{6}\left(\lambda_i^0-\lambda_{i-1}^0\right) & 0 \\ 0 & \sigma^2\left(\frac{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}{\Gamma^2(1-d_i^0)(1-2d_i^0)}+A_i^{\mu}\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)\right) \end{pmatrix}$$

with $A_i^{\mu}\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)$ defined in (1.7). Because of the term $A_i^{\mu}\left(\lambda_{i-1}^0,\lambda_i^0,d_i^0\right)$,

$$A\left(\lambda_i^0,\lambda_{i-1}^0,\theta_i^0\right)\neq D\left(\lambda_i^0,\lambda_{i-1}^0,\theta_i^0\right).$$

Proof. Part a) Write

$$
\begin{aligned}
S_{i,T}\left(\lambda_{i-1},\lambda_i,\theta_i\right) &= \sum_{t=T_{i-1}+1}^{T_i}\left(\Delta_{t-T_{i-1}}^{d_i}\Delta_t^{-d_i}u_t\right)^2+\sum_{t=T_{i-1}+1}^{T_i}\left(\left(\mu_i-\mu_i^0\right)\Delta_{t-T_{i-1}}^{d_i}1\right)^2 \\
&\quad -2\sum_{t=T_{i-1}+1}^{T_i}\Delta_{t-T_{i-1}}^{d_i}\Delta_t^{-d_i}u_t\left(\mu_i-\mu_i^0\right)\Delta_{t-T_{i-1}}^{d_i}1.
\end{aligned}
$$

For the first term uniformly in d_i and μ_i,

$$
\frac{1}{T}\sum_{t=T_{i-1}+1}^{T_i}\left(\Delta_{t-T_{i-1}}^{d_i}\Delta_t^{-d_i}u_t\right)^2\xrightarrow{p}\left(\lambda_i^0-\lambda_{i-1}^0\right)\sum_{j=0}^{\infty}\pi_j^2(d_i-d_i^0),
$$

a limit that has a unique minimum at d_i^0. The convergence follows from a law of large numbers and the last expression follows from (19) in Lobato and Velasco (2007). Uniformity, follows from a similar argument as the one in the proof of Lemma 13. For the second term uniformly in d_i and μ_i,

$$
T^{2d_i-1}\sum_{t=T_{i-1}+1}^{T_i}\left(\left(\mu_i-\mu_i^0\right)\Delta_{t-T_{i-1}}^{d_i}1\right)^2\to\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}\frac{\left(\mu_i^0-\mu_i\right)^2}{\left(1-2d_i^0\right)\Gamma^2\left(1-d_i^0\right)},
$$

a limit that has a unique minimum at $\mu_i=\mu_i^0$. Uniformity follows from the deterministic character. Finally, the third term multiplied by T^{d_i-1} is uniformly in d_i and μ_i of order $o_p(1)$.

Part b)

The derivative evaluated at true break points and true parameters, $F_t\left(\lambda_{i-1}^0,\theta_i^0\right)$, for $t=T_{i-1}^0+1,...,T_i^0$,

$$
F_t\left(\lambda_{i-1}^0,\theta_i^0\right)=\left(\begin{array}{c}+\sum_{j=1}^{t-T_{i-1}^0-1}j^{-1}u_{t-j}+\dot{\Delta}_{t-T_{i-1}^0}^{d_i^0}\sum_{j=t-T_{i-1}^0}^{t-1}\pi_j\left(-d_i^0\right)u_{t-j} \\ -\Delta_{t-T_{i-1}^0}^{d_i}1\end{array}\right). \tag{1.28}
$$

First, the (1,1) element of $D_T\left(\theta_i^0\right)$ multiplied by T^{-1} converges in mean square to $\left(\lambda_i^0-\lambda_{i-1}^0\right)\frac{\pi^2}{6}$ because the terms coming from the second term in $F_t\left(\lambda_{i-1}^0,\theta_i^0\right)$ are negligible. The (2,2) element of $D_T\left(\theta_i^0\right)$ multiplied by $T^{-1+2d_i^0}$ converges to $\frac{\left(\lambda_i^0-\lambda_{i-1}^0\right)^{1-2d_i^0}}{\Gamma^2\left(1-d_i^0\right)\left(1-2d_i^0\right)}$. Finally, the (1,2) element is of smaller order.

Part c) Note that for a break fraction λ_{i-1}, the residuals for $t=T_{i-1}+1,...,T_i^0$ are

$$
\hat{u}_t^{(i)}\left(\lambda_{i-1},\theta_i\right)=\Delta_{t-T_{i-1}}^{d_i}\left(\mu_i^0-\mu_i\right)+\Delta_{t-T_{i-1}}^{d_i-d_i^0}u_t+\Delta_{t-T_{i-1}}^{d_i}\sum_{j=t-T_{i-1}}^{t-1}\pi_j\left(-d_i^0\right)u_{t-j}.
$$

The difficulty arises from showing that the last term is asymptotically negligible. Similarly, the derivatives $F_t(\lambda, \theta)$ have a similar additional term. For the (2,2) element of $D_{i,T}\left(\theta, \theta_i^0\right)$, *fidi* convergence corresponds to the one in part b) since $\left(s - \lambda_{i-1}^0\right) = O_p\left(T^{-1}\right)$. Uniformity follows directly from the fact that the term is deterministic. For $D_{i,T}\left(\theta, \theta_i^0\right)_{(1,1)}$, we use that terms containing $\left(\mu_i^0 - \mu_i\right)$ are of order $T^{1-2d_i^0}$. For uniformity, in (s, r, θ), the tightness of $D_T\left(\theta, \theta_i^0\right)$ can be proved using Johansen and Nielsen (2010).

Part d) The $(1x1)$ element of $A_i\left(\theta_i^0\right)$ is straightforward. For the $(2x2)$ element, we separate the second term into two uncorrelated terms

$$\Delta_{t-\lambda_{i-1}^0 T}^{d_i} \Delta_t^{-d_i^0} u_t = \Delta_{t-\lambda_{i-1}^0 T}^{d_i} \Delta_{t-\lambda_{i-1}^0 T}^{-d_i^0} u_t + \Delta_{t-\lambda_{i-1}^0 T}^{d_i} \sum_{k=t-T_{i-1}^0}^{t-1} \pi_k\left(-d_i^0\right) u_{t-k}.$$

The first term leads to a variance component of $\frac{\left(\lambda_i^0 - \lambda_{i-1}^0\right)^{1-2d_i^0}}{\Gamma\left(1-d_i^0\right)\left(1-2d_i^0\right)}$. The one corresponding to the second term,

$$\begin{aligned} & Var\left(T^{d_i-1/2} \sum_{t=T_{i-1}^0+1}^{T_i^0} \Delta_{t-T_{i-1}^0}^{d_i} 1 \Delta_{t-T_{i-1}^0}^{d_i} \sum_{k=t-T_{i-1}^0}^{t-1} \pi_k\left(-d_i^0\right) u_{t-k}\right) \\ = \; & T^{2d_i-1} E\left[\sum_{k=1}^{T_{i-1}^0}\left(\sum_{t=1}^{T_i^0-T_{i-1}^0} \pi_{t-1}\left(d_i-1\right) \Delta_t^{d_i} \pi_{T_{i-1}^0+t-k}\left(-d_i^0\right)\right) u_k\right]^2, \end{aligned}$$

converges to $\sigma^2 A_i^\mu\left(\lambda_{i-1}^0, \lambda_i^0, \theta_i^0\right)$. Combining the two terms leads to the result. ■

Lemma 16 discusses the estimators for the partitions (T_1, T_2, T_3), (T_1, T_2^0, T_3) and (T_1, T_2, T_2^0, T_3).

Lemma 16 *(Behavior of estimators)*

a) For the estimator $(\theta_2^{***}, \theta_2^\delta, \theta_3^{***})$ for (T_1, T_2, T_2^0, T_3)

$$\begin{aligned} \left(d_2^{***} - d_2^0, d_2^\delta - d_2^0, d_3^{***} - d_3^0\right) &= \left(O_p(T^{-1/2}), O_p(\Delta_2^{-1/2}), O_p(T^{-1/2})\right) \\ \left(\mu_2^{***} - \mu_2^0, \mu_2^\delta - \mu_2^0, \mu_3^{***} - \mu_3^0\right) &= \left(O_p(T^{-1/2+d_2^0}), O_p(\Delta_2^{-1/2+d_2^0}), O_p(T^{-1/2+d_3^0})\right) \end{aligned}$$

b) For the estimator $(\theta_2^*, \theta_3^{**})$ for (T_1, T_2, T_3)

$$\begin{aligned} \left(d_2^* - d_2^0, d_3^{**} - d_3^0\right) &= \left(O_p(T^{-1/2}), O_p(T^{-1/2})\right) \\ \left(\mu_2^* - \mu_2^0, \mu_3^{**} - \mu_3^0\right) &= \left(O_p(T^{-1/2+d_2^0}), O_p(T^{-1/2+d_3^0})\right) \end{aligned}$$

c) For the estimator $(\theta_2^{**}, \theta_3^*)$ for (T_1, T_2^0, T_3)

$$\begin{aligned} \left(d_2^{**} - d_2^0, d_3^* - d_3^0\right) &= \left(O_p(T^{-1/2}), O_p(T^{-1/2})\right) \\ \left(\mu_2^{**} - \mu_2^0, \mu_3^* - \mu_3^0\right) &= \left(O_p(T^{-1/2+d_2^0}), O_p(T^{-1/2+d_3^0})\right) \end{aligned}$$

Lemmata 17 and 18 are needed for the proof of Theorem 2. We analyze the terms $\sum d_t^2(s,\theta_i)$, $\sum d_t(s,\theta)\, u_t\left(s,\theta^0\right)$ multiplied by Δ_2^{-1} in the case of breaks in memory and mean or only in memory and by $\Delta_2^{-1+2d_2^\delta}$ in the case of a break only in the mean respectively. Both Lemmata use Lemma 16. The proofs of tightness are similar to the ones of Lemma 13 and use among others Lemma 15 and 16 of Johansen and Nielsen (2010). Further, we consider $T_2 < T_2^0$ and $\left(s - \lambda_1^0\right) = O_p\left(T^{-1}\right)$.

Lemma 17 *(Break in memory or in memory and mean.)*
a) *Behavior of* $\sum d_t^2$. *For* $r = \lambda_2 < \lambda_2^0$

$$\Delta_2^{-1} \sum_{t=[sT]+1}^{[rT]} d_t^2 = o_p(1), \Delta_2^{-1} \sum_{t=\lambda_2^0 T+1}^{[\lambda_3 T]} d_t^2 = o_p(1).$$

$$\Delta_2^{-1} \sum_{t=[rT]+1}^{\lambda_2^0 T} d_t^2 \xrightarrow{p} \sum_{j=1}^{\infty} \pi_j^2 \left(d_2^\delta - d_2^0\right) = O_p(1),$$

b) *Behavior of* $\sum d_t u_t$.

$$\Delta_2^{-1} \sum_{t=[sT]+1}^{[rT]} d_t u_t, \sum_{t=[rT]+1}^{\lambda_2^0 T} d_t u_t, \Delta_2^{-1} \sum_{t=\lambda_2^0 T+1}^{\lambda_3 T} d_t u_t = o_p(1)$$

Proof. We use Cauchy Schwarz for the first and third in Part b). In particular

$$\left[\Delta_2^{-1} \sum_{t=[sT]+1}^{[rT]} d_t u_t\right]^2 \leq \Delta_2^{-1} \sum_{t=[sT]+1}^{[rT]} d_t^2 \Delta_2^{-1} \sum_{t=\lambda_1^0 T+1}^{[rT]} u_t^2,$$

where the first term converges to zero from Part a). The proofs are similar to the one of Lemma 1 with the difference that the considered interval is constant rather than proportional to T. In particular, some tedious analysis shows that the terms converge uniformly. ∎

Lemma 18 *For* $(d_2^\delta - d^0) = O_p(\Delta_2^{-1/2})$.

a) *Behavior of* $\sum d_t^2$

$$\Delta_2^{-1+2d_2^\delta} \sum_{t=[sT]+1}^{[rT]} d_t^2 = o_p(1), \Delta_2^{-1+2d_2^\delta} \sum_{t=\lambda_2^0 T+1}^{\lambda_3 T} d_t^2 = o_p(1)$$

$$\Delta_1^{-1+2d_2^\delta} \sum_{t=[rT]+1}^{\lambda_2^0 T} d_t^2 \xrightarrow{p} \frac{\left(\mu_2^0 - \mu_2^\delta\right)^2}{\Gamma^2(1-d^0)(1-2d^0)}$$

b) *Behavior of* $\sum d_t u_t$

$$\begin{aligned}
\Delta_2^{-1+2d_2^\delta} \sum_{t=[sT]+1}^{[rT]} d_t u_t &= o_p(1), \Delta_2^{-1+2d_2^\delta} \sum_{t=[rT]+1}^{\lambda_2^0 T} d_t u_t = o_p(1) \text{ and} \\
\Delta_2^{-1+2d_2^\delta} \sum_{t=\lambda_2^0 T+1}^{\lambda_3 T} d_t u_t &= o_p(1)
\end{aligned}$$

Proof. The terms including μ are deterministic, for the terms including d we can show that they converge uniformly at a faster rate and are, therefore, negligible at the present rate. Part b) follows from similar argument as the one in Part a). In addition we need also a uniform argument for the terms including μ. ■

Propositions

Proposition 19 derives the asymptotic distribution of the estimators defined below (1.9) under the local alternative $H_{1,T}$.

Proposition 19 *Under Assumptions 1-2, for* $i = 1, ..., k+1$

$$a) \qquad \bar{T}\left(d_1^0\right)\left(\hat{\theta}_{1,i} - \theta_1^0\right) \Longrightarrow \left(\begin{array}{c} (\lambda_i)^{-1}\left(\frac{\sqrt{6}}{\pi} B^h(\lambda_i)\right) \\ \frac{\sigma}{\lambda_i^{1-2d_1^0}}\left(\Gamma\left(1-d_1^0\right)\sqrt{1-2d_1^0}\tilde{W}^h(\lambda_i)\right) \end{array}\right).$$

$$b) \qquad \bar{T}\left(d_1^0\right)\left(\hat{\theta}_i - \theta_1^0\right) \Longrightarrow$$

$$\left(\begin{array}{c} (\lambda_i - \lambda_{i-1})^{-1}\left(\frac{\sqrt{6}}{\pi}\left[B^h(\lambda_i) - B^h(\lambda_{i-1})\right]\right) \\ \frac{1}{\lambda_i^{1-2d_1^0} - \lambda_{i-1}^{1-2d_1^0}}\left(\sigma\Gamma\left(1-d_1^0\right)\sqrt{1-2d_1^0}\left[\tilde{W}^h(\lambda_i) - \tilde{W}^h(\lambda_{i-1})\right]\right) \end{array}\right),$$

where $B^h(\lambda_i)$ *and* $\tilde{W}^h(\lambda)$ *are defined in (1.13) and (1.14) respectively.* θ_i *and* θ_j *are asymptotically uncorrelated.*

Proof. Part a) The consistency follows from combining Lemma 3a) and Robinson and Hualde (2010). For the asymptotic distribution of the estimator, we analyze its denominator and numerator. For the denominator, we obtain uniformly,

$$\bar{T}\left(d_1^0\right)\sum_{1,i} F_t\left(0,\theta_{1,i}\right)F_t'\left(0,\theta_{1,i}\right)\bar{T}\left(d_1^0\right)\xrightarrow{p}\left(\begin{array}{cc}\lambda_i\frac{\pi^2}{6} & 0\\ 0 & \frac{\lambda_i^{1-2d_1^0}}{\Gamma^2\left(1-d_1^0\right)\left(1-2d_1^0\right)}\end{array}\right)$$

and for the numerator, we obtain

$$\bar{T}\left(d_1^0\right)\sum_{1,i} u_t F_t\left(0,\theta_{1,i}\right)\Longrightarrow\left(\begin{array}{c}\frac{\pi}{\sqrt{6}}B^h\left(\lambda_i\right)\\ \frac{1}{\Gamma\left(1-d_1^0\right)\sqrt{1-2d_1^0}}\tilde{W}^h\left(\lambda_i\right)\end{array}\right),$$

where the weak convergence to Brownian and fractional Brownian Motion follows from a FCLT and Marinucci and Robinson (1999) respectively. The fractional Brownian Motion $\tilde{W}_{1/2-d_1^0}\left(\lambda_i\right)$ has the same marginal distribution as the standard one $W_{1/2-d_1^0}\left(\lambda_i\right)=\int_0^{\lambda_i}\left(\lambda_i-r\right)dB\left(r\right)$. Because of the opposite order of summing the error terms, its covariance is (1.12) rather than the usual one,

$$\begin{aligned}Cov\left(W_{1/2-d_1^0}\left(\lambda_i\right),W_{1/2-d_1^0}\left(\lambda_{i-1}\right)\right) &= \frac{\lambda_i^{1-2d_1^0}+\lambda_{i-1}^{1-2d_1^0}}{\Gamma\left(1-d_1^0\right)\left(1-2d_1^0\right)}\\ &\quad -E\left[W_{1/2-d_1^0}\left(\lambda_i\right)-W_{1/2-d_1^0}\left(\lambda_{i-1}\right)\right]^2.\end{aligned}$$

In consequence, $\tilde{W}_{1/2-d_1^0}\left(.\right)$ has independent increments. The local drift of the memory estimator

$$\frac{\frac{1}{T}\sum\limits_{j=1}^{i+1}\sum\limits_{t=T_{j-1}+1}^{T_j}h_d\left(\frac{t}{T}\right)\left(\sum\limits_{j=1}^{t-1}\dot{\pi}_j(0)u_{t-j}\right)^2}{T^{-1}\sum\limits_{j=1}^{i+1}\sum\limits_{t=T_{j-1}+1}^{T_j}\left(\sum\limits_{j=1}^{t-1}\dot{\pi}_j(0)u_{t-j}\right)^2}\xrightarrow{p}\frac{\sigma^2}{\lambda_i}\int_0^{\lambda_i}h_d\left(u\right)du.$$

and the one of the mean

$$\frac{T^{2d_1^0-1}\sum\limits_{j=1}^{i+1}\sum\limits_{t=T_{j-1}+1}^{T_j}(\Delta_t^{d_{1,i}}1)\Delta_t^{d_{1,i}}h_\mu\left(\frac{t}{T}\right)}{T^{2d_1^0-1}\sum\limits_{j=1}^{i+1}\sum\limits_{t=T_{j-1}+1}^{T_j}(\Delta_t^{d_{1,i}})^2}\xrightarrow{p}\frac{\Gamma^2\left(1-d_1^0\right)\left(1-d_1^0\right)D\left(\lambda_i,d_1^0,h_\mu\right)}{\lambda_i^{1-2d_1^0}},$$

where for the denominator we use that $(\Delta_t^{d_i}1)\simeq(t-1)^{-d_i}$ and where the numerator converges to

$$D\left(\lambda_i,d_1^0,h_\mu\right)=\lim_{T\to\infty}T^{2d_1^0-1}\sum_{t=1}^{[\lambda_i T]}\pi_{t-1}\left(d_1^0-1\right)\sum_{j=0}^{t-1}\pi_{j-1}\left(d_1^0\right)h_\mu\left(\frac{t-j}{T}\right).$$

Part b) The proofs follow similar lines as the one of Part a). The variance of the estimator μ_i is $\frac{\lambda_i^{1-2d_1^0}-\lambda_{i-1}^{1-2d_1^0}}{\Gamma^2\left(1-d_1^0\right)\left(1-2d_1^0\right)}$. Further, the covariance of the two estimators μ_i and μ_j for $i<j$ is

$$Cov\left(T^{d_1^0-1/2}\left(\mu_i-\mu_i^0\right),T^{d_1^0-1/2}\left(\mu_j-\mu_j^0\right)\right)=0,$$

since unlike Lemma 15,

$$Cov\left(T^{1/2-d_1^0}\sum_{t=T_{i-1}+1}^{T_i}F_t\left(0,\theta_{1,i}\right)u_t,T^{d_1^0-1/2}\sum_{t=T_{j-1}+1}^{T_j}F_t\left(0,\theta_{1,i}\right)u_t\right)=0.$$

Thus, the estimator using the filter (1.9) is uncorrelated under H_0 which contrasts the one in Theorem 4. ■

For ℓ vs. $\ell+1$ breaks, Proposition 20 derives the asymptotic distribution of the unconstrained estimators for the i's regime, assuming *one* additional break in this regime. Let $\tau=\hat{T}_{i-1}+\gamma(\hat{T}_i-\hat{T}_{i-1})$ be the additional break point in regime i.

Proposition 20 *Under Assumptions 1-3 for $i=1,...,\ell+1$ and under H_0^ℓ :*

$$a)\ \bar{T}\left(d_i^0\right)\left(\hat{\theta}_{i,\tau}-\theta_i^0\right)\Longrightarrow\left(\begin{array}{c}\frac{\sqrt{6}}{\pi}B^h\left(\gamma\right)/\gamma\\ \frac{\sigma\sqrt{1-2d_1^0}\Gamma\left(1-d_1^0\right)\hat{W}^h(\gamma)}{\gamma^{1-2d_i^0}}\end{array}\right).$$

$$b)\ \bar{T}\left(d_i^0\right)\left(\hat{\theta}_{\tau,i+1}-\theta_i^0\right)\Longrightarrow\left(\begin{array}{c}\frac{\sqrt{6}}{\pi}B^h\left(1-\gamma\right)/\left(1-\gamma\right)\\ \frac{\sigma\sqrt{1-2d_i^0}\Gamma\left(1-d_i^0\right)\left(\hat{W}^h(1)-\hat{W}^h(\gamma)\right)}{1-\gamma^{1-2d_i^0}}\end{array}\right).$$

Proof. Part a) The behavior of the denominator of the estimator follows from Lemma 15. First, the l break fractions are superconsistently estimated. We can use arguments similar to the ones in Theorem 3, to show for the numerator

$$\bar{T}\left(d_i^0\right)\sum_{t=T_{i-1}^0+1}^{\tau}u_t\left(\lambda_{i-1}^0,\theta_i^0\right)F_t\left(\lambda_{i-1}^0,\theta_i^0\right)\Rightarrow\left(\begin{array}{c}\frac{\pi}{\sqrt{6}}B\left(\gamma\left(\lambda_i^0-\lambda_{i-1}^0\right)\right)\\ \frac{\sigma\tilde{W}_{1/2-d_i^0}\left(\gamma\left(\lambda_i^0-\lambda_{i-1}^0\right)\right)}{\sqrt{1-2d_i^0}\Gamma\left(1-d_i^0\right)}+C\left(\lambda_{i-1}^0,\gamma,d_i^0\right)\end{array}\right)$$

where $C\left(\lambda_{i-1}^0,\gamma,d_i^0\right)$ is discussed in (1.22). In particular, the convergence of the first component follows from a functional central limit theorem. For the convergence of the second component, we use Marinucci and Robinson (1999) and that (1.21) converges in distribution to $C\left(\lambda_{i-1}^0,\gamma,d_i^0\right)$. The additional term is again a consequence of the too short filter.

Part b) follows similarly. ■

Proposition 21 analyzes the estimators corresponding to the ones in Propositions 19 in the bootstrap world. $\overset{p}{\Longrightarrow}$ denotes weak convergence in probability as defined in Gine and Zinn (1990).

Proposition 21 *Under Assumptions 1 and 2 and under H_0 or $H_{1,T}$, the estimators $\hat{\theta}^*$ and $\hat{\theta}^*_{1,i}$ converge weakly in probability ($\overset{p}{\Longrightarrow}$) to the same limits as the ones in Propositions 19.*

Proof. The proof follows from combining results about the convergence of partial sums in the bootstrap world to fractional Brownian Motions with the behavior of the estimators in Propositions 19. It remains to show these convergence results. For this, we incorporate into Kapetanios' (2010) analysis, the estimation of the mean but for a process without a short memory component. Since we analyze the behavior of the bootstrap under $H_0/H_{1,T}$, we filter under the assumption of no breaks. In the notation of Kapetanios (2010), we have to show his Theorem 1

$$\tilde{W}^*_{T,1/2-\hat{d}} = \frac{1}{T^{\hat{d}-1/2}} \sum_{t=1}^{[rT]} \pi_{t-1}\left(\hat{d}-1\right) u_t^* \Longrightarrow \tilde{W}_{1/2-d_1^0}(r) \text{ in probability,}$$

where the convergence is in the sense of Giné and Zinn (1990). $\tilde{W}_{1/2-d_1^0}(r)$ is the fractional Brownian Motion of order $1/2-d_1^0$ defined in Proposition 19, and u_t^* is a bootstrap resample of the residuals of the regression under SSR_0. Hence Kapetanios' (2010) first assumption is clearly satisfied. We have to show

$$\begin{aligned} &1)\ E^*|u_t^*|^r < \infty \text{ in probability for some } r > 2. \\ &2) \sup_r |\tilde{W}^*_{T,1/2-d^0}(r) - \tilde{W}^*_{T,1/2-\hat{d}}(r)| = o_{p*}(1). \end{aligned}$$

For 1), we have to show that

$$\frac{1}{T}\sum_{t=1}^{T} |\hat{u}_t - \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t|^r = O_p(1)$$

Write

$$\frac{1}{T}\sum_{t=1}^{T} |\hat{u}_t - \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t|^r \leq c\left(A_T + D_T + E_T\right)$$

where

$$A_T = \frac{1}{T}\sum_{t=1}^{T}|u_t|^r, D_T = |\frac{1}{T}\sum_{t=1}^{T}u_t|^r \leq KA_T \text{ and } E_T = \frac{1}{T}\sum_{t=1}^{T}|\hat{u}_t - u_t|^r.$$

First, as in Park (2002), A_T and D_T are of order $O_p(1)$. Consider E_T

$$\frac{1}{T}\sum_{t=1}^{T}\left|\hat{u}_t - u_t\right|^r = \frac{1}{T}\sum_{t=1}^{T}\left|\Delta_t^d\left(\mu_1^0 + T^{d_1^0-1/2}h_\mu\left(\frac{t}{T}\right) - \mu\right) + \sum_{j=1}^{t-1}\pi_j\left(d - d_1^0 - T^{-1/2}h_d\left(\frac{t}{T}\right)\right)u_{t-j}\right|^r,$$

where the second term is $o_p(1)$ following from eq. (4.17) in Wright (1995) and the fact that h_d is bounded. Using $(\hat{\mu} - \mu_1^0) = O(T^{d_1^0-1/2})$ and the boundedness of h_μ, the first term is also of order $o_p(1)$.

For 2), we need to show

$$\max_s \frac{1}{T^{d_1^0-1/2}}\left|\sum_{t=1}^{s}\pi_{t-1}\left(\hat{d} - 1\right)u_t^* - \sum_{t=1}^{s}\pi_t\left(d_1^0 - 1\right)u_t^*\right| = o_{p*}(1)$$

where u_t^* is an *iid* heterogenous process in the bootstrap probability space, drawn with probability $1/T$ from the residuals $\hat{u}_t$. In particular, defining $v_j^* = u_{t-j}^*, j = 1, ..., t$, the proof follows the same steps as the one in Kapetanios (2010).

Similarly, partial sums converge to Brownian Motions. ■

1.11 Appendix B: Proofs

Proof of Proposition 1

Part a) We show that the memory estimation is still consistent for $d^0 > 0$, but inconsistent for $d^0 \leq 0$. We analyze heuristically the case of inconsistent estimation of μ with $0 < d^0 < 1/2$. In particular, for $|\hat{\mu} - \mu^0| > C$ the objective function

$$\frac{1}{T}SSR = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\left[\left(\mu^0 - \mu\right)\Delta_t^d 1 + \Delta_t^{d-d^0}u_t\right]^2 \tag{1.29}$$

converges uniformly in $d \in D$ and μ to

$$\sum_{j=1}^{\infty}\pi_j^2(d - d^0).$$

Therefore, the SSR is still minimized at the true parameter d^0 if $0 < d^0 < 1/2$. The asymptotically negligible terms

$$\left(\mu^0 - \mu\right)^2 K_1 T^{-2d} + \frac{1}{T}2\left(\mu^0 - \mu\right)\sum_{t=1}^{T}\pi_t(d-1)\sum_{j=1}^{t-1}\pi_j(d - d^0)u_{t-j}$$

lead to the mentioned finite sample effects which depend on $d^0, (\mu^0 - \mu)$ and T. Especially, for d^0 close to 0, the bias can be huge leading to a highly upward biased estimator in finite samples. On the support $0 \leq d < 1/2$, the limit of the expression (1.29) is not continuous due to the additional term $I(d=0)(\mu^0 - \mu)^2$. Clearly, for $|\hat{\mu} - \mu^0| > C$, (1.29) is in the limit not minimized in $d = 0$. In consequence, the estimator is not consistent for $d^0 = 0$. The same argument is obviously true if we do not estimate μ, just set $\hat{\mu} = 0$.

Part b) We have to show that

$$\hat{\mu}(d) - \mu^0 = O_p\left(T^{d^0 - 1/2}\right) \text{ uniformly in } d \in D,$$

by showing convergence of the *fidi* and tightness. For tightness we show in supplemental Appendix that

$$E|T^{1/2-d^0}\left(\hat{\mu}(d_2) - \mu^0\right) - T^{1/2-d^0}\left(\hat{\mu}(d_1) - \mu^0\right)|^2 \leq K|d_2 - d_1|^2. \tag{1.30}$$

Proof of Theorem 2

We provide the main steps of the proof and indicate where they differ from the ones of Boldea and Hall (2010). Define

$$d_t(\lambda_{k-1}, \theta_k) = \hat{u}_t(\lambda_{k-1}, \theta_k) - u_t\left(\lambda_{k-1}, \theta_k^0\right), \tag{1.31}$$

where $\hat{u}_t(\lambda_{k-1}, \theta_k)$ is defined in (1.4), for $t \in I_j^0 \cap \hat{I}_k$ with $I_j^0 = [T_{j-1}^0 + 1, T_j^0]$ and $\hat{I}_k = [\hat{T}_{k-1} + 1, \hat{T}_k]$ and $k, j = 1, ..., m+1$. $d_t(\lambda_{k-1}, \theta_k)$ and $\hat{u}_t^{(k)}(\lambda_{k-1}, \theta_k)$ depend also on $\left\{\theta_i^0\right\}, \left\{\theta_{i-1}^0, \theta_i^0\right\}$ and $\left\{\theta_{i-1}^0, \theta_i^0, \theta_{i+1}^0\right\}$ in the cases $\lambda_{k-1}^0 < \lambda_{k-1} < t < \lambda_k^0, \lambda_{k-1} < \lambda_{k-1}^0 < t < \lambda_k^0$ and $\lambda_{k-1} < \lambda_{k-1}^0 < \lambda_k^0 < t$ respectively. Boldea and Hall (2010) work with a different expression separating true quantities from estimated ones. In our case, both are fractionally integrated and we work rather with expression (1.31). First, we focus on the break in T_i^0. For simplicity, we denote $d_t(\lambda_{k-1}, \theta_k)$ and $\hat{u}_t(\lambda_{k-1}, \theta_k)$ as d_t and $\hat{u}_t$. From the CSS estimation we get

$$\sum_{t=1}^{T} \hat{u}_t^2 = \sum_{t=1}^{T} u_t^2\left(\lambda_{k-1}, \theta_k^0\right) + \sum_{t=1}^{T} d_t^2 + 2\sum_{t=1}^{T} d_t u_t\left(\lambda_{k-1}, \theta_k^0\right)$$

implying that

$$T^{-\delta_i}\sum_{t=1}^{T} d_t^2 + 2T^{-\delta_i}\sum_{t=1}^{T} d_t u_t\left(\lambda_{k-1}, \theta_k^0\right) \leq 0, \tag{1.32}$$

where $\delta_i = 1$ for a break in T_i^0 in memory and mean or only in memory and $\delta_i = 1 - 2d_i^0$ for a break only in the mean. Denoting $q_T \sim O_p\left(T^b\right)$ if $P\left(|q_T| > T^b\right) < \bar{\eta}$ for $T \geq T(\bar{\eta})$ for some $b\epsilon R$ and any $\bar{\eta} > 0$ and $q_T \sim O_p^+\left(T^b\right)$ if $plim q_T$ is positive, the proof of the consistency works by showing that $T^{-\delta_i}\sum_{t=1}^{T} d_t u_t = o_p(1)$ and $T^{-\delta_i}\sum_{t=1}^{T} d_t^2 = O_p^+(1)$, when the break fraction λ_i is inconsistently estimated. In particular, we use Lemma 13 and 14 for proving Theorem 34. Inequality (1.32) together with Part (i) of Lemma 14 would imply that $T^{-\delta_i}\sum_{t=1}^{T} d_t^2 = o_p(1)$ which would contradict part (ii) of Lemma 14. In particular, Lemma 14 is also true for an estimator $\hat{\lambda}_i < \lambda_i^0$ and, in consequence, the break fraction is not estimated too low. The same argument applies for $\hat{\lambda}_i > \lambda_i^0$ and we conclude that the break fraction estimator is consistent.

Proof of Theorem 3

This proof follows the proof of Theorem 2 of Boldea and Hall (2010). We consider the case of *three* breaks.We analyze two different cases of changing parameters that require a different analysis:

- case A: a break in memory and mean or a break in memory.
- case B: a break in mean; $d_1^0 = d_2^0 = d_3^0 \geq 0$.

Consistency of the three breaks is already established. Because of consistency we only have to consider the behavior of the break points in

$$V_{\in} = \left\{ (T_1, T_2, T_3) : |T_i - T_i^0| \leq \varepsilon T \ (i = 1, 2, 3) \right\}.$$

First, consider case $\hat{T}_2 < T_2^0$. In contrast to Boldea and Hall (2010), here the argument is not symmetric and we have to consider also the case $T_2^0 > T_2$. The proof works basically by showing that the break point is with a very small probability in the set

$$V_{\in}(C) = \left\{ (T_1, T_2, T_3) : |T_i - T_i^0| \leq \varepsilon T \ (i = 1, 2, 3);\ \Delta_2 = T_2^0 - T_2 > C \right\}.$$

Hence with large probability $|\hat{T}_2 - T_2^0| < C$. We will show that if $T_2 \in V_{\in}(C)$,

$$P\left\{ \min_{V_{\in}(C)} \frac{S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)}{\Delta_2^{\delta}} \leq 0 \right\} < \eta, \text{ for } T \geq T(\eta), \tag{1.33}$$

contradicting the sum of squares minimization and implying that T_2 does not belong to $V_{\in}(C)$. For case A, $\delta = 1$ and, for case B, $\delta = 1 - 2d_2^0$. Define

$$\begin{aligned} SSR1 &= S_T(T_1, T_2, T_3),\ SSR2 = S_T\left(T_1, T_2^0, T_3\right) \text{ and} \\ SSR3 &= S_T\left(T_1, T_2, T_2^0, T_3\right). \end{aligned}$$

We show that

$$S_T(T_1, T_2, T_3) - S_T\left(T_1, T_2^0, T_3\right) = (SSR1 - SSR3) - (SSR2 - SSR3)$$

is positive with high probability for large T picking ε and C. The behavior of the corresponding estimators is discussed in Lemma 16. We locate the dominating terms in $S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)$ and show that at least some are positive with large probability. Equation (1.47) is equivalent to

$$\Delta_2^{-\delta}\left(SSR1 - SSR2\right) \sim O_p^{+}\left(1\right) \tag{1.34}$$

We introduce some notation:

$$I_1 = [1, T_1]; I_2 = [T_1 + 1, T_2], I_2^{\Delta} = [T_2 + 1, T_2^0],\ I_3 = [T_2^0 + 1, T_3],\ I_4 = [T_3 + 1, T].$$

Next,

$$\begin{aligned} \frac{SSR_1 - SSR_3}{\Delta_2^{\delta}} &= \frac{1}{\Delta_2^{\delta}}\left[\sum\nolimits_{I_2^{\Delta}}\left[u_t^2\left(\theta_3^{**}\right) - u_t^2\left(\theta_2^{\delta}\right)\right] + \sum\nolimits_{I_3}\left[u_t^2\left(\theta_3^{**}\right) - u_t^2\left(\theta_3^{***}\right)\right]\right] \\ &= D_1 + D_2. \end{aligned}$$

Since θ_3^{**} estimates θ_3^0 and θ_2^{δ} estimates θ_2^0, there is a mismatch in D_1, while there is none in D_2 (θ_3^{**} and θ_3^{*} estimate θ_3^0). In the supplemental Appendix, we use Lemmata 17 and 18 to show in a similar way as in Boldea and Hall (2010) that D_1 dominates in the limit D_2. We further show that $\Delta_2^{-\delta}\left(SSR2 - SSR3\right) \sim o_p\left(1\right)$.

In Theorem 1 and 2, we focus on the break in T_i and assume that all other break fractions are estimated consistently (Theorem 34) and at the rate T (Theorem 36). It suffices to discuss this case since it is the least favorable case for the contradiction that is used for deriving the consistency of the break fraction λ_i.

Proof of Theorem 4

We first obtain consistency and $\sqrt{T}$-rate convergence of the estimator d_i when it is calculated with estimated rather than true endpoints. Given these results, we establish $T^{1/2-d_i^0}$ rate convergence for the estimator of μ_i. Finally, we show that the estimators using the estimated break points have the same asymptotic distribution as the ones using the true ones.

We start with the asymptotic distribution of the estimators assuming that the break points are the true ones. By the superconsistency of the break fractions, this distribution will correspond to the one when the break points are estimated. First, the consistency of the estimator d_i follows from Lemma 15a). The asymptotic distribution of the estimator follows from Lemma 15a) and b). Because the residuals evaluated at the true parameters and true break fractions $u_t\left(\lambda_{i-1}^0, \theta_i^0\right)$ differ from u_t, the variance of the mean estimator contains the additional term (1.6). Similarly, the covariance between the estimators μ_i and μ_j

$$D_{ij}^{\mu}\left(\left\{\lambda_k^0, \lambda_{k-1}^0, d_k^0\right\}_{k=i,j}\right) = \frac{\Gamma^2\left(1-d_i^0\right)\left(1-2d_i^0\right)}{\left(\lambda_i^0 - \lambda_{i-1}^0\right)^{1-2d_i^0}}\frac{\Gamma^2\left(1-d_j^0\right)\left(1-2d_j^0\right)}{\left(\lambda_j^0 - \lambda_{j-1}^0\right)^{1-2d_j^0}}A_{ij}^{\mu} \tag{1.35}$$

where

$$A_{ij}^{\mu} = \lim_{T\to\infty} T^{-1} \sum_{k=1}^{\left[\lambda_{i-1}^{0}T\right]} \left(T^{d_i^0} \sum_{t=1}^{\left[\left(\lambda_i^0-\lambda_{i-1}^0\right)T\right]} \pi_{t-1}\left(d_i^0-1\right) \sum_{j=0}^{t} \pi_j\left(d_i^0\right) \pi_{\left[\lambda_{i-1}^0 T\right]+t-j-k}\left(-d_i^0\right) \right)$$
$$\cdot \left(T^{d_j^0} \sum_{t=1}^{T_j^0-T_{j-1}^0} \pi_{t-1}\left(d_j^0-1\right) \sum_{j=0}^{t} \pi_j\left(d_j^0\right) \pi_{T_{j-1}^0+t-j-k}\left(-d_j^0\right) \right).$$

Consequently, the estimators μ_i and μ_j are not asymptotically independent.

Next, the proof of consistency of the parameter estimates in the two regimes using the estimated rather than the true break points and the proof that the asymptotic distribution corresponds to the one assuming the true break point follows the same lines as in Boldea and Hall (2010).

Proof of Theorem 5

For deriving the asymptotic distribution of the test statistic (1.10), we, first, show for the denominator under the local alternative:

$$\begin{aligned} SSR_k(\boldsymbol{\lambda}) &= \sum_{i=1}^{k+1} \frac{1}{T} \sum_{t=T_{i-1}+1}^{T_i} \left(\Delta_t^{d_i} 1 \left(\mu_1^0 - T^{d_1^0-1/2} h_\mu\left(\frac{t}{T}\right) - \mu_i \right) + \Delta_t^{d_i-d_1^0-\frac{1}{\sqrt{T}}h_d\left(\frac{t}{T}\right)} u_t \right)^2 \\ &= \lambda_1 \sum_{k=0}^{\infty} \pi_k^2\left(d_1-d_1^0\right) + \sum_{i=2}^{k+1} \left(\lambda_i-\lambda_{i-1}\right) \sum_{k=0}^{\infty} \pi_k^2\left(d_i-d_1^0\right) + o_p(1) = \sigma^2 + o_p(1) \end{aligned}$$

where the terms including $(\mu_j^0-\mu_i)$ are negligible by Lemma 13 and the convergence is a consequence of Lemma 15 a). Next, we discuss the behavior of the numerator. As in Boldea and Hall (2010), we write

$$SSR_0 - SSR_k(\boldsymbol{\lambda}) = \sum_{t=1}^{T} u_t^2(\hat{\theta}) - \sum_{i=1}^{k+1} \sum_{t=\lambda_{i-1}T+1}^{\lambda_i T} u_t^2(\hat{\theta}_i) = ... = \sum_{i=1}^{k} F_{T,i}^{*}$$

with

$$F_{T,i}^{*} = D^R(1,i+1) - D^R(1,i) - D^U(i+1,i+1) \tag{1.36}$$

where the index $1,i$ indicates summing over $[1,T_i]$ and i over $[T_{i-1}+1,T_i]$. $D^R(1,i) = \sum_{1,i}[u_t^2(\hat{\theta}_{1,i}) - u_t^2]$ and $D^U(i,i) = \sum_i [u_t^2(\hat{\theta}_i) - u_t^2]$. We start with the term $D^R(1,i)$

$$D^R(1,i) = \sum\nolimits_{1,i} d_t^2(\hat{\theta}_{1,i},\theta_1^0) - 2\sum\nolimits_{1,i} u_t d_t(\hat{\theta}_{1,i},\theta_1^0) = I^R + II^R$$

As in Boldea and Hall (2010), using a mean value theorem (MVT),

$$\begin{aligned} I^R &= \left[T^{1/2}(\hat{d}_{1,i}-d_1^0)\right]^2 T^{-1}\sum_{1,i} F_{d,t}^2\left(\bar{\theta}_{1,i,t}\right) \\ &\quad+\left[T^{1/2-d_1^0}\left(\hat{\mu}_{1,i}-\mu_1^0\right)\right]^2 T^{-1+2d_1^0}\sum_{1,i} F_{\mu,t}^2\left(\bar{\theta}_{1,i,t}\right) \\ II^R &= 2\left[T^{1/2}(\hat{d}_{1,i}-d_1^0)\right] T^{-1/2}\sum_{1,i} u_t F_{d,t}\left(\bar{\theta}_{1,i,t}\right) \\ &\quad+2\left[T^{1/2-d_1^0}\left(\hat{\mu}_{1,i}-\mu_1^0\right)\right] T^{-1/2+d_1^0}\sum_{1,i} u_t F'_{\mu,t}\left(\bar{\theta}_{1,i,t}\right) \end{aligned}$$

where $\bar{\theta}_{1,i,t}$ lies in the segment line $\hat{\theta}_{1,i}$ and θ_1^0. Also here since $\bar{\theta}_{1,i,t} \xrightarrow{p} \theta_1^0$ for each t and $E\left[F_t(\theta) F_t'(\theta)\right]$ has uniform bounds, from Proposition 19 part b) and its proof, we obtain

$$D^R(1,i) \Longrightarrow -\sigma^2\left(\lambda_i^{-1}\left(B^h(\lambda_i)\right)^2+\lambda_i^{-1+2d_1^0}\left(\tilde{W}^h(\lambda_i)\right)^2\right) \tag{1.37}$$

For the term $D^U(i,i)$ using Proposition 19 and similar arguments as the previous ones, we obtain,

$$\begin{aligned} D^U(i,i) \Longrightarrow\ & -\sigma^2\left[(\lambda_i-\lambda_{i-1})^{-1}\left(B^h(\lambda_i)-B^h(\lambda_{i-1})\right)^2\right. \\ & \left.+\left(\lambda_i^{1-2d_1^0}-\lambda_{i-1}^{1-2d_1^0}\right)^{-1}\left(\tilde{W}^h(\lambda_i)-\tilde{W}^h(\lambda_{i-1})\right)^2\right]. \end{aligned}$$

Finally, combining the two terms and using a continuous mapping theorem (CMT) for the *sup* functional leads to the stated test statistic.

The independence of the estimates of memory and mean, discussed in Theorem 4, implies the additiveness of the test statistic.

Proof of Theorem 7

First, the estimated break fractions converge to the true ones at rate T, for breaks in the memory H_1^d, the mean H_1^μ and in both $\mathrm{H}_1^{d,\mu}$. Under the alternative, the test statistic (1.10) diverges since its denominator still converges to σ^2 because break fractions and regime parameters are consistently estimated. If there is at least one break in the memory or in memory and mean, $D^R(1,i)$ is of order $O_p(T)$ and $D^U(i,i)$ is of order $O_p(T^{1-2d_i^0})$ because the mean estimators stop being consistent. Thus, the test statistic diverges at rate T. Equally, we find that, if only the mean is changing, $SSR_0 - SSR_k(\boldsymbol{\lambda}) = O_p(T^{1-2d^0})$ and the test statistic diverges at rate T^{1-2d^0}. If we tested for a break only in the memory or only in the mean, the tests reject under the alternative of a break in the tested and in both parameters. Under the alternative of a break only in the not tested parameter, the tests reject asymptotically with probability α.

Proof of Proposition 9

Under the hypothesis of one break at λ_1^0, the estimator λ_1 converges at rate T to the break fraction λ_1^0.

Proof of a)

Components from Theorem 5 involving the estimation of the mean are negligible. Finally, the components involving the memory behave as in Theorem 5 with the difference that now λ_1 does not have a spurious limit and thus the limit will be a function of the true break fraction. Therefore, the test statistic corresponds to the one of a usual Chow test.

Proof of b)

For testing a break in the mean, terms involving the break in the memory are again negligible. Using the filter truncated at the supposed break points, we obtain for the estimator of the mean,

$$\left(\hat{\mu} - \mu^0\right) = \frac{\sum_{t=1}^{\lambda_1 T} \left(\Delta_t^{d_1} 1\right) \hat{u}_t + \sum_{t=\lambda_1 T+1}^{T} \left(\Delta_{t-\lambda_1 T}^{d_2} 1\right) \hat{u}_t}{\sum_{t=1}^{\lambda_1 T} \left(\Delta_t^{d_1} 1\right)^2 + \sum_{t=\lambda_1 T+1}^{T} \left(\Delta_{t-\lambda_1 T}^{d_2} 1\right)^2}.$$

It is easy to show that for $d_1^0 < d_2^0$ in numerator and denominator, the first term dominates and for $d_1^0 > d_2^0$ the second one does. In (1.36), the first and third term cancel. From the second term of (1.36), follows the result. For the latter, as mentioned before, $\hat{u}_t$ contains some term similar to the one in Theorem 7 coming from a too short filter causing the increased variance.

Proof of Theorem 8

Under $\mathrm{H}_0 : m = l$, as in Boldea and Hall (2010), the test statistic can be written as

$$F_T\left(l+1|l\right) = \max_{1\leq i\leq l} \sup_{\tau\in\Delta_{i,\eta}} F_{T,i}\left(l+1|l\right)/\hat{\sigma}_i^2$$

where $F_{T,i}\left(l+1|l\right) = SSR(\hat{T}_{i-1}, \hat{T}_i) - SSR(\hat{T}_{i-1}, \tau) - SSR(\tau, \hat{T}_i)$ with $SSR(\hat{T}_{i-1}, \hat{T}_i)$ being the sum of squared residuals for the segment $[\hat{T}_{i-1}, \hat{T}_i]$. Based on Proposition 20, the proof follows using similar arguments to the ones in Theorem 5.

Proof of Theorem 10

We show that the bootstrap based test (1.25) has the same asymptotic distribution as the one in Theorem 5. The estimates $\hat{d}, \hat{\mu}$ play the role of the true parameter values in Theorem 5. The estimates $\hat{\theta}_{1,i}^*, \hat{\theta}_i^*$ denote the estimates for the bootstrap data $\{y_t^*\}_{t=1}^T$. From Proposition 21, these estimators converge weakly in probability to the same limits as the ones in Proposition 19.

For establishing the asymptotic distribution of the test statistic, first we have to show for the denominator that

$$SSR_k^*\left(\boldsymbol{\lambda}\right)/\left(T-\left(k+1\right)p\right) = \sigma^2 + o_p^*.$$

In particular,

$$\begin{aligned} SSR_k^*(\boldsymbol{\lambda}) &= \sum_{i=1}^{k+1} \frac{1}{T} \sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \left((\hat{\mu} - \hat{\mu}_i^*) \Delta_t^{\hat{d}_i^*} 1 + \Delta_t^{\hat{d}_i^* - \hat{d}} u_t \right)^2 \\ &= \sum_{i=1}^{k+1} (\lambda_i - \lambda_{i-1}) \sum_{k=0}^{\infty} \pi_k^2 \left(\hat{d}_i^* - \hat{d} \right) + o_p^*(1) = \sigma^2 + o_p^*(1) \end{aligned}$$

To prove the convergence we show that $E^*\left[\frac{1}{T} SSR_1^*\right] = \hat{\sigma}^2$ and $Var^*\left[\frac{1}{T} SSR_1^*\right] = o_p(1)$. For the former,

$$E^* \sum_{i=1}^{k+1} \frac{1}{T} \sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \left((\hat{\mu} - \hat{\mu}_i^*) \Delta_t^{\hat{d}_i^*} 1 + \Delta_t^{\hat{d}_i^* - \hat{d}} u_t \right)^2 = \hat{\sigma}^2 \frac{1}{T} \sum_{i=1}^{k+1} \sum_{t=[\lambda_{i-1}T]+1}^{[\lambda_i T]} \sum_{j=1}^{t-1} \pi_j^2 \left(\hat{d}_i^* - \hat{d} \right).$$

For the second term we apply a variant of the Lemma 1, substituting d_1^0 by $\hat{d}$, and a similar argument as the one for the first term. The convergence follows from $T \to \infty$ and the fact that $\hat{d}_1^*$ and $\hat{d}$ converge to d_1^0 and $\hat{\sigma}^2$ converges to σ^2. The behavior of the numerator follows from applying Proposition 21 to the Proof of Theorem 5. Finally, from applying a CMT, we obtain that $\sup_\lambda F_T^*(\lambda, k, p)$ converges weakly in probability to the corresponding limit in Theorem 5 for $h_d = h_\mu = 0$.

Part c)

The test is consistent because the bootstrap test statistic converges to a constant and the original test statistic diverges. For the former, under H_1, the estimators $\hat{d}$ and $\hat{\mu}$ converge to weighted averages of the true parameter values. Applying the test to the newly integrated series, the resulting test statistic has still a bounded limit distribution. Since, from Theorem 7, the test statistic diverges under H_1, the bootstrap test is consistent.

Proof of Proposition 11

We first show 1). Note that terms including μ are uniformly of order $o_p(1)$. For $i = 1$,

$$T^{1/2}\left(d_1(\lambda) - d^0\right) = \frac{T^{-1/2} \sum_{t=1}^{[\lambda T]} \left(\sum_{j=0}^{t-1} \dot{\pi}_j(0) u_{t-j} \right) u_t}{T^{-1} \sum_{t=1}^{[\lambda T]} \left(\sum_{j=0}^{t-1} \dot{\pi}_j(0) u_{t-j} \right)^2} + o_p(1).$$

For $j = 1, 2$, N_j denotes the numerator and D_j the denominator of $d_1(\lambda_j)$. Thus,

$$d_1(\lambda_2) - d_1(\lambda_1) = \frac{N_2}{D_2} - \frac{N_1}{D_1} = ... = \frac{N_1}{D_1 D_2}(D_2 - D_1) + \frac{1}{D_2}(N_1 - N_2).$$

In consequence, we can show tightness for numerator and denominator separately. For the latter, for showing tightness we need to show that

$$\left\| T^{-1} \sum_{t=[\lambda_1 T]+1}^{[\lambda_2 T]} \left(\sum_{j=0}^{t-1} \dot{\pi}_j (0) u_{t-j} \right)^2 \right\|_2^2 \leq K \left|\lambda_2 - \lambda_1\right|^2,$$

From a triangle inequality,

$$\left\| T^{-1} \sum_{t=[\lambda_1 T]+1}^{[\lambda_2 T]} \left(\sum_{j=0}^{t-1} \dot{\pi}_j (0) u_{t-j} \right)^2 \right\|_2 \leq T^{-1} \sum_{t=[\lambda_1 T]+1}^{[\lambda_2 T]} \left\| \left(\sum_{j=0}^{t-1} \dot{\pi}_j (0) u_{t-j} \right)^2 \right\|_2 \leq K \left|\lambda_2 - \lambda_1\right|,$$

where the boundedness of the norm follows from previous arguments.

$$T^{-1} \sum_{t=1}^{[\lambda T]} \left(\sum_{j=0}^{t-1} \dot{\pi}_j (0) \right)^2$$

For the numerator, weak convergence in λ follows from a standard FCLT.

Next using $d_i (\lambda) - d^0 = O_p \left(T^{-1/2}\right)$, we show 2). For $i = 1$,

$$T^{1/2-d^0} \left(\mu_1 (\lambda) - \mu^0\right) = \frac{T^{d^0-1/2} \sum_{t=1}^{[\lambda T]} \left(\Delta_t^{d^0} 1\right) u_t}{T^{2d^0-1} \sum_{t=1}^{[\lambda T]} \left(\Delta_t^{d^0} 1\right)^2} + o_p (1).$$

Next,

$$\mu_1 (\lambda_2) - \mu_1 (\lambda_1) = \frac{N_1}{D_1 D_2} (D_2 - D_1) + \frac{1}{D_2} (N_1 - N_2).$$

Tightness for the denominator follows directly from its deterministic character. For the numerator, we can apply a fractional FCLT (Marinucci and Robinson, 1999) to show that it converges weakly to a fractional Brownian Motion.

1.12 Supplemental Appendix

In this Supplemental Appendix, I provide the proofs for Lemma 13, Lemma 15, Proposition 1 Part a) and Theorem 3.

Proof of Lemma 13

We have to show uniform convergence of the terms $\sum_{t=[sT]+1}^{[rT]} d_t^2$ and $\sum_{t=[sT]+1}^{[rT]} d_t u_t$. In principal, we have to consider four cases i) $\lambda_{i-1}^0 \leq s < r \leq \lambda_i^0$, ii) $s \leq \lambda_{i-1}^0 < r \leq \lambda_i^0$, iii) $\lambda_{i-1}^0 \leq s \leq \lambda_i^0 < r$ and iv) $s \leq \lambda_{i-1}^0 \leq \lambda_i^0 < r$. Being the most involved case, we focus in the following on case d). All other cases follow from similar but simpler arguments. The following processes converge uniformly in r and θ and in s for $\left(s - \lambda_{i-1}^0\right) = O\left(T^{-1}\right)$. We provide the proof of Part a).

For the case of breaks in both parameters or a break only in the memory, we show that for a generic $(s, r, \theta), s \leq \lambda_{i-1}^0 \leq \lambda_i^0 < r$, uniformly in r and θ

$$T^{-1} \sum_{t=[sT]+1}^{[rT]} d_t^2 \xrightarrow{p} \left(r - \lambda_i^0\right) \sigma^2 \sum_{j=1}^{\infty} \pi_j^2 (d - d_{i+1}^0) + \left(\lambda_i^0 - \lambda_{i-1}^0\right) \sigma^2 \sum_{j=1}^{\infty} \pi_j^2 (d - d_i^0)$$

In particular, note that for $t = T_{i-1} + 1, ..., T_i^0$,

$$d_t\left(\lambda_{i-1}, \theta_i\right) = \left(\mu_i^0 - \mu_i\right) \Delta_{t-T_{i-1}}^{d_i} 1 + \left(\Delta_{t-T_{i-1}}^{d_i - d_i^0} - 1\right) u_t + \Delta_{t-T_{i-1}}^{d_i} \sum_{j=t-T_{i-1}}^{t-1} \pi_j\left(-d_i^0\right) u_{t-j},$$

and for $t = T_i^0 + 1, ..., T_i$,

$$\begin{aligned} d_t\left(\lambda_{i-1}, \theta_i\right) &= \Delta_{t-T_i^0}^{d_{i+1}}\left(\mu_{i+1}^0 - \mu_i\right) + \left(\Delta_{t-T_{i-1}}^{d_i} - \Delta_{t-T_i^0}^{d_i}\right)\left(\mu_i^0 - \mu_i\right) \\ &\quad + \left(\Delta_{t-T_i^0}^{d_i - d_{i+1}^0} - 1\right) u_t + \Delta_{t-T_i^0}^{d_i} \sum_{j=t-T_i^0}^{t-1} \pi_j\left(-d_{i+1}^0\right) u_{t-j} \\ &\quad + \left(\Delta_{t-T_{i-1}}^{d_i} - \Delta_{t-T_i^0}^{d_i}\right) \Delta_t^{-d_i^0} u_t. \end{aligned}$$

Next, we write

$$T^{-1} \sum_{t=[sT]+1}^{[rT]} d_t^2 = T^{-1} \sum_{t=[sT]+1}^{\lambda_{i-1}^0 T} d_t^2 + T^{-1} \sum_{t=\lambda_{i-1}^0 T+1}^{\lambda_i^0 T} d_t^2 + T^{-1} \sum_{t=\lambda_i^0 T+1}^{[rT]} d_t^2.$$

We analyze the third term which involves both parameters s and r. The other two terms follow from similar arguments. First,

$$T^{-1}\sum_{t=\lambda_i^0T+1}^{[rT]} d_t^2 = \sum_{t=\lambda_i^0T+1}^{[rT]}\Big[\Delta_{t-\lambda_i^0T}^{d_i}\left(\mu_{i+1}^0-\mu_i\right)+\left(\Delta_{t-\lambda_{i-1}^0T}^{d_i}-\Delta_{t-\lambda_i^0T}^{d_i}\right)\left(\mu_i^0-\mu_i\right)$$
$$+\left(\Delta_{t-[sT]}^{d_i}-\Delta_{t-\lambda_{i-1}^0T}^{d_i}\right)\left(\mu_{i-1}^0-\mu_i\right)$$
$$+\left(\Delta_{t-\lambda_i^0T}^{d_i-d_{i+1}^0}-1\right)u_t+\Delta_{t-\lambda_i^0T}^{d_i}\sum_{j=t-[sT]}^{t-1}\pi_j\left(-d_{i+1}^0\right)u_{t-j}$$
$$+\left(\Delta_{t-\lambda_{i-1}^0T}^{d_i}-\Delta_{t-\lambda_i^0T}^{d_i}\right)\Delta_t^{-d_i^0}u_t+\left(\Delta_{t-[sT]}^{d_i}-\Delta_{t-\lambda_{i-1}^0T}^{d_i}\right)\Delta_t^{-d_{i-1}^0}u_t\Big]^2.$$

This can be written as

$$T^{-1}\sum_{t=\lambda_i^0T+1}^{[rT]}A^2+T^{-1}\sum_{t=\lambda_i^0T+1}^{[rT]}B^2+2T^{-1}\sum_{t=\lambda_i^0T+1}^{[rT]}AB \tag{1.38}$$

where

$$A=\left[\begin{array}{c}\Delta_{t-\lambda_i^0T}^{d_i}\left(\mu_{i+1}^0-\mu_i\right)+\left(\Delta_{t-\lambda_{i-1}^0T}^{d_i}-\Delta_{t-\lambda_i^0T}^{d_i}\right)\left(\mu_i^0-\mu_i\right)\\ +\left(\Delta_{t-[sT]}^{d_i}-\Delta_{t-\lambda_{i-1}^0T}^{d_i}\right)\left(\mu_{i-1}^0-\mu_i\right)\end{array}\right]$$

and

$$B=\left[\begin{array}{c}+\left(\Delta_{t-\lambda_i^0T}^{d_i-d_{i+1}^0}-1\right)u_t+\Delta_{t-\lambda_i^0T}^{d_i}\sum_{j=t-[sT]}^{t-1}\pi_j\left(-d_{i+1}^0\right)u_{t-j}\\ +\left(\Delta_{t-\lambda_{i-1}^0T}^{d_i}-\Delta_{t-\lambda_i^0T}^{d_i}\right)\Delta_t^{-d_i^0}u_t+\left(\Delta_{t-[sT]}^{d_i}-\Delta_{t-\lambda_{i-1}^0T}^{d_i}\right)\Delta_t^{-d_{i-1}^0}u_t\end{array}\right]$$

The first term of (1.38) is of order $o_p(1)$, uniformly in the parameters, because it is deterministic and bounded. Next, the second term of (1.38) has expectation

$$\left(r-\lambda_i^0\right)\sigma^2\sum_{j=1}^{\infty}\pi_j^2(d-d_{i+1}^0).$$

coming from the first term of B. The other terms can be written as

$$\sum_{t=\lambda_i^0T+1}^{[rT]}\left[\sum_{j=t-\lambda_i^0T}^{t-1}\psi_j u_{t-j}\right]^2$$

with uncorrelated errors and where for $j<\lambda_i^0T+1$, $\psi_j=\pi_j(d-d_{i+1}^0)$ and for $j>\lambda_i^0T+1$, ψ_j is differently defined. Its expectation equals

$$\sigma^2\sum_{t=\lambda_i^0T+1}^{[rT]}\left[\sum_{j=t-\lambda_i^0T}^{t-1}\psi_j^2\right].$$

This expectation can be shown to be *zero*. Since it is a sum of positive terms with an expectation of zero, we obtain mean square convergence. Further, the variance of the first component converges again to zero. This establishes the *fidi* convergence. Further, the process has to be tight in (s, r, θ). For Tightness, we use Lemma 15 and 16 of Johansen and Nielsen (2009). In particular, we write the second term as

$$A\left(d,s,r\right)=T^{-1}\sum_{t=\lambda_i^0T+1}^{[rT]}\left(\sum_{j=1}^{t-1}\psi_j u_{t-j}\right)^2.$$

For that we have to show for some $d_1 < d < d_2$, some $r_1 < r < r_2$ and some $s_1 < s < s_2$

$$\begin{aligned}
& E\left\{\left|A\left(d_2,r_2,s_2\right)-A\left(d,r,s\right)\right|^4\left|A\left(d_1,r_1,s_1\right)-A\left(d,r,s\right)\right|^4\right\}\leq\left|\left(d_2,r_2,s_2\right)-\left(d_1,r_1,s_1\right)\right|^4 \\
\Leftrightarrow\quad & ||AB||_4\leq\left|\left(d_2,r_2,s_2\right)-\left(d_1,r_1,s_1\right)\right|
\end{aligned}$$

First, we can show for the RHS,

$$\begin{aligned}
K\left|\left(d_2,r_2,s_2\right)-\left(d_1,r_1,s_1\right)\right| &= K\sqrt{\left(d_2-d_1\right)^2+\left(r_2-r_1\right)^2+\left(s_2-s_1\right)^2} \\
&> K_1\left(d_2-d_1\right)+K_2\left(r_2-r_1\right)+K_3\left(s_2-s_1\right)
\end{aligned}$$

Next, the LHS,

$$\begin{aligned}
& ||\ \left|A\left(d_2,r_2,s_2\right)-A\left(d,r_2,s_2\right)+A\left(d,r_2,s_2\right)-A\left(d,r_2,s\right)+A\left(d,r_2,s\right)-A\left(d,r,s\right)\right| \\
& \cdot\left|A\left(d_1,r_1,s_1\right)-A\left(d,r_1,s_1\right)+A\left(d,r_1,s_1\right)-A\left(d,r_1,s\right)+A\left(d,r_1,s\right)-A\left(d,r,s\right)\right|\ ||_4
\end{aligned}$$

can be written as a sum consisting of terms of a change in d such as

$$|\,|[A\left(d_2,r_2,s_2\right)-A\left(d,r_2,s_2\right)]\left[A\left(d_1,r_1,s_1\right)-A\left(d,r_1,s_1\right)\right]|\,|_4,$$

terms of a change in r

$$||\left[A\left(d,r_2,s\right)-A\left(d,r,s\right)\right]\left[A\left(d,r_1,s\right)-A\left(d,r,s\right)\right]|\,|_4,$$

terms of a change in s

$$||\left[A\left(d,r_2,s_2\right)-A\left(d,r_2,s\right)\right]\left[A\left(d,r_1,s_1\right)-A\left(d,r_1,s\right)\right]|\,|_4,$$

and crossterms of them. First, we analyze the change in d,

$$
\begin{aligned}
& \left\| \left[T^{-1} \sum_{t=\lambda_i^0 T+1}^{[rT]} \left(\sum_{j=1}^{t-1} \psi_j\left(d_2\right) u_{t-j} \right)^2 - \left(\sum_{j=1}^{t-1} \psi_j\left(d\right) u_{t-j} \right)^2 \right] \cdot \right. \\
& \left. \left[T^{-1} \sum_{t=\lambda_i^0 T+1}^{[rT]} \left(\sum_{j=1}^{t-1} \psi_j(d) u_{t-j} \right)^2 - \left(\sum_{j=1}^{t-1} \psi_j(d_0) u_{t-j} \right)^2 \right] \right\|_4 \\
\leq \;\; & \left| d_2 - d_1 \right| .
\end{aligned}
$$

Next, by the Cauchy Schwarz inequality, the LHS $\leq ||A||_8 ||B||_8$. From applying a Taylor approximation,

$$
\begin{aligned}
||A||_8 &= \left\| T^{-1} \sum_{t=\lambda_i^0 T+1}^{[rT]} \left(\sum_{j=1}^{t-1} \psi_j(d_2) u_{t-j} \right)^2 - \left(\sum_{j=1}^{t-1} \psi_j(d) u_{t-j} \right)^2 \right\|_8 \\
&\simeq \left\| T^{-1} \sum_{t=\lambda_i^0 T+1}^{[rT]} 2 \left(\sum_{j=1}^{t-1} \psi_j(d^*) u_{t-j} \right) \left(\sum_{j=1}^{t-1} \dot{\psi}_j(d^*) u_{t-j} \right) |d_2 - d_1| \right\|_8 \qquad (1.39)
\end{aligned}
$$

with $d < d^* < d_2$. Now we can use Lemma 16 of Johansen and Nielsen (2009) with

$$
\begin{aligned}
X_t &= \sum_{j=1}^{t-1} \psi_j(d^*) u_{t-j} \text{ with } \sum_{j=1}^{t-1} \psi_j^2(d^*) < \infty \\
U_t &= \sum_{j=1}^{t-1} \dot{\psi}_j(d^*) u_{t-j} \text{ with } \sum_{j=1}^{t-1} \dot{\psi}_j^2(d^*) < \infty
\end{aligned}
$$

Hence, by term (1.39)

$$
\begin{aligned}
\leq \;\; & c|d_2 - d_1| \frac{1}{T} \sum_{t=\lambda_i^0 T+1}^{[rT]} \left\| 2 \left(\sum_{j=1}^{t-1} \psi_j(d^*) u_{t-j} \right) \right\|_2 \left\| \left(\sum_{j=1}^{t-1} \dot{\psi}_j(d^*) u_{t-j} \right) \right\|_2 \\
\leq \;\; & K|d_2 - d_1|.
\end{aligned}
$$

Second, we analyze the term for a change in r. Similar to before,

$$
||A||_8 = \left\| \frac{1}{T} \sum_{t=[rT]+1}^{[r_2 T]} \left(\sum\nolimits_{j=1}^{t-1} \psi_j\left(d_2\right) u_{t-j} \right)^2 \right\|_8
$$

with $X_t = U_t = \sum_{j=1}^{t-1} \psi_j\left(d_2\right) u_{t-j}$. Hence, by Lemma 16 of Johansen and Nielsen (2009), this term is bounded by

$$K\frac{1}{T}\sum_{t=[rT]+1}^{r_2T}\left\|\left(\sum_{j=1}^{t-1}\psi_j\left(d_2\right)u_{t-j}\right)^2\right\|_2 \leq K|r_2-r|$$

Similarly, we can show that the terms of a change in s, for $s-\lambda_{i-1}^0 = O\left(T^{-1}\right)$are bounded by $K|s_2-s_1|$.

Finally, we obtain by the Cauchy Schwarz inequality for the cross terms

$$\begin{aligned} & \|\left[A\left(d_1,r_1,s_1\right)-A\left(d,r_1,s_1\right)\right]\left[A\left(d,r_2,s_2\right)-A\left(d,r,s\right)\right]_4\| \\ \leq \quad & \|\left[A\left(d_1,r_1,s_1\right)-A\left(d,r_1,s_1\right)\right]||_8||\left[A\left(d,r_2,s_2\right)-A\left(d,r,s\right)\right]\|_8 \end{aligned}$$

. But, $|\left[A\left(d_1,r_1,s_1\right)-A\left(d,r_1,s_1\right)\right]||_8 \leq K|d_2-d_1|$ (shown before) establishes the result. The term $\left(\lambda_{i-1}^0-s\right)\sigma^2\sum_{j=1}^{\infty}\pi_j^2(d-d_{i-1}^0)$ vanishes uniformly in s for $\left(s-\lambda_{i-1}^0\right)=O_p\left(T^{-1}\right)$.

For the Part b), we have to show uniformly in r and θ

$$T^{-1}\sum_{t=[sT]+1}^{[rT]}d_tu_t = o_p\left(1\right).$$

In particular,

$$T^{-1}\sum_{t=[sT]+1}^{[rT]}d_tu_t = T^{-1}\sum_{t=[sT]+1}^{\lambda_{i-1}^0T}d_tu_t + T^{-1}\sum_{t=\lambda_{i-1}^0T+1}^{\lambda_i^0T}d_tu_t + T^{-1}\sum_{t=\lambda_i^0T+1}^{[rT]}d_tu_t.$$

where we focus again on the last term,

$$T^{-1}\sum_{t=\lambda_i^0T+1}^{[rT]}d_tu_t$$

This mean zero process is a MDS since d_t is orthogonal to u_t and has a variance of

$$\sigma^2T^{-2}\sum_{t=\lambda_i^0T+1}^{[rT]}d_t^2$$

which is from Part a) of order $o_p\left(1\right)$. Tightness follows again from Johansen and Nielsen and a similar argument as the previous one.

Next, we analyze the case of breaks only in the mean. First, for the Part a), we have to show for $(d_i - d^0) = O\left(T^{-1/2}\right)$ and $\left(s - \lambda_{i-1}^0\right) = O_p\left(T^{-1}\right)$, uniformly in θ, r,

$$T^{-1+2d} \sum_{t=[sT]+1}^{[rT]} d_t^2 \xrightarrow{p} \left(\mu_{i+1}^0 - \mu\right)^2 \frac{\left(r - \lambda_i^0\right)^{1-2d^0}}{\Gamma^2\left(1-d^0\right)\left(1-2d^0\right)} + \left(\mu_i^0 - \mu\right)^2 \frac{\left(\lambda_i^0 - \lambda_{i-1}^0\right)^{1-2d^0}}{\Gamma^2\left(1-d^0\right)\left(1-2d^0\right)}.$$

In particular, the terms including memory terms are negligible because all involved memory parameters are consistently estimated at rate $T^{1/2}$. Since the terms including μ are deterministic, the tightness follows straightforwardly. In general, the proofs are comparable to the ones of Lemma 3.

Finally for the Part b), we have to show, uniformly in θ, r,

$$T^{-1+2d} \sum_{t=[sT]+1}^{[rT]} d_t u_t = o_p(1).$$

In particular, the process is again a MDS with mean zero and variance of order $o_p(1)$ implying convergence of the FIDI. The tightness argument follows again from Johansen and Nielsen and from a similar argument as the one in the previous proof.

Proof of Lemma 15

Part c) Define $F_t(\theta) = \frac{\partial f_t(\theta)}{\partial \theta}$, a $px1$ vector, a function of θ_i for $t \in [T_{i-1}+1, T_i]$ and $F_{k,t}(\theta)$, $k = d, \mu$ the derivative with respect to d and μ respectively. The proofs correspond to the proofs of Lemma 1 substituting when necessary π by $\dot{\pi}$. Further, we have to deal with the fact that $u_t\left(\lambda_{i-1}^0, \theta_i^0\right) \neq u_t$. However, for the latter $u_t\left(\lambda_{i-1}^0, \theta_i^0\right) - u_t$ can be written as $\sum_{j=1}^{\lambda_{i-1}^0 T} \phi_j u_j$, a term uncorrelated with F_t and of smaller order. Finally, since the terms stay summable, the proofs go through after some modifications. The resulting residuals are for $t = T_{i-1}, ..., T_i^0$,

$$\hat{u}_t^{(i)}\left(\lambda_{i-1}, \theta_i\right) = \Delta_{t-T_{i-1}}^{d_i}\left(\mu_i^0 - \mu_i\right) + \Delta_{t-T_{i-1}}^{d_i - d_i^0} u_t + \Delta_{t-T_{i-1}}^{d_i} \sum_{j=t-T_{i-1}}^{t-1} \pi_j\left(-d_i^0\right) u_{t-j}$$

and the difficulty arises from showing that the last term of the first is asymptotically negligible. Similarly, the derivatives $F_t(\lambda, \theta)$ have a similar additional term. First, the $(2x2)$ element corresponds to the one in Part b) since $\left(s - \lambda_{i-1}^0\right) = O_p\left(T^{-1}\right)$

$$D_{i,T}\left(\theta, \theta_i^0\right)_{(2,2)} = T^{2d-1} \sum_{t=[sT]+1}^{[rT]} \left(\Delta_{t-[sT]}^d 1\right)^2 \xrightarrow{p} \frac{\left(r - \lambda_{i-1}^0\right)^{1-2d}}{\Gamma^2(1-d)(1-2d)}$$

Uniformity follows directly from the fact that the term is deterministic. Next, it can be shown for the element $(1x1)$

$$
\begin{aligned}
D_{i,T}\left(\theta,\theta_i^0\right)_{(1,1)} &= T^{-1}\sum_{t=[sT]+1}^{[rT]} F_{dt}^2\left(s,\theta\right) = T^{-1}\sum_{t=\lambda_i^0 T+1}^{[rT]} F_{dt}^2\left(s,\theta\right) + T^{-1}\sum_{t=[sT]+1}^{\lambda_i^0 T} F_{dt}^2\left(s,\theta\right) \\
&\xrightarrow{p} \left(r-\lambda_i^0\right)\sigma^2\sum_{k=1}^{\infty}\dot{\pi}_k^2(d-d_{i+1}^0) + \left(\lambda_i^0-\lambda_{i-1}^0\right)\sigma^2\sum_{k=1}^{\infty}\dot{\pi}_k^2(d-d_i^0),
\end{aligned}
$$

where $F_{dt}\left(s,\theta\right)$ is the derivative with respect to the memory. In particular, terms containing $\left(\mu_i^0-\mu_i\right)$ are of order $T^{1-2d_i^0}$. Further, we drop again some negligible terms, and convergence follows from a similar argument as the one in the proof of Lemma 1. For uniformity, in (s,r,θ), the tightness of $D_T\left(\theta,\theta_i^0\right)$ can be proved in a similar way as it is done in the proof of Lemma 1. Finally, it can be shown that

$$
D_{i,T}\left(\theta,\theta_i^0\right)_{(1,2)} = o_p\left(1\right).
$$

Proof of Lemma 16

a) The estimator $\left(d_2^{***},d_2^{\delta},d_3^{***},\mu_2^{***},\mu_2^{\delta},\mu_3^{***}\right)$ minimizes

$$
\begin{aligned}
&\sum_{t=T_1}^{T_3}\hat{u}_t^2 - \sum_{t=T_1}^{T_3}u_t^2 \\
=\ &\sum_{t=T_1}^{T_2}d_t^2 + \sum_{t=T_2+1}^{T_2^0}d_t^2 + \sum_{t=T_2^0+1}^{T_3}d_t^2 - \sum_{t=T_1}^{T_2}d_tu_t - \sum_{t=T_2+1}^{T_2^0}d_tu_t - \sum_{t=T_2^0+1}^{T_3}d_tu_t.
\end{aligned}
$$

Using T-rate convergence of λ, established in Theorem 2, and as a consequence of Lemma 1, the objective function converges when multiplied by $1/T$ uniformly to

$$
\left(\lambda_2^0-\lambda_1^0\right)\sum_{j=1}^{\infty}\pi_j\left(d_2^{***}-d_2^0\right) + \left(\lambda_3^0-\lambda_2^0\right)\sum_{j=1}^{\infty}\pi_j\left(d_3^{***}-d_3^0\right),
$$

which is minimized in $d_2^{***}=d_2^0$ and $d_3^{***}=d_3^0$ for any $\left(d_2^{\delta},\mu_2^{***},\mu_2^{\delta},\mu_3^{***}\right)$. Hence, d_2^{***} and d_3^{***} are consistent at rate $T^{1/2}$. Next given this $T^{1/2}$ consistent estimators d_2^{***} and d_3^{***}, from Lemma 1, the objective function converges, when multiplied by $T^{2d_2^0-1}$, uniformly to

$$
\frac{\left(\mu_2^0-\mu_2^{***}\right)^2\left(\lambda_2^0-\lambda_1^0\right)^{1-2d_2^0}}{\Gamma^2\left(1-d_2^0\right)\left(1-2d_2^0\right)},
$$

which is minimized at $\mu_2^{***}=\mu_2^0$. We assume without loss of generality $d_2^0<d_3^0$. Hence, we have established $T^{1/2-d_2^0}$ rate convergence of μ_2^{***}. Using this convergence, the objective

function when multiplied by $T^{2d_3^0-1}$ converges uniformly to

$$\frac{\left(\mu_3^0-\mu_3^{***}\right)^2\left(\lambda_3^0-\lambda_2^0\right)^{1-2d_3^0}}{\Gamma^2\left(1-d_3^0\right)\left(1-2d_3^0\right)},$$

which is minimized at $\mu_3^{***}=\mu_3^0$, establishing the $T^{1/2-d_3^0}$ rate convergence of μ_3^{***}. Both limits follow from Lemma 1. Using these convergence results, the objective function converges uniformly at rate Δ_2, again as a consequence of Lemma 5, to

$$\left(\lambda_2^0-\lambda_2\right)\sum_{j=1}^{\infty}\pi_j\left(d_2^\delta-d_2^0\right),$$

which is minimized in true $d_2^\delta=d_2^0$. Finally, we can show from Lemma 6 that the objective funtion after substituting all convergence results, multiplied by $\Delta_2^{2d_2^0-1}$ converges uniformly to

$$\frac{\left(\mu_2^0-\mu_2^\delta\right)^2}{\Gamma^2\left(1-d\right)\left(1-2d\right)}.$$

This limit is clearly minimized in $\mu_2^\delta=\mu_2^0$. Hence, we have established $\Delta_2^{1/2-d_2^0}$ convergence of the estimator.

The proofs of **b)** and **c)** follow from similar but simpler arguments.

Proof of Proposition 1

Part a) We show that

$$\hat{\mu}\left(d\right)-\mu^0=O_p\left(T^{d^0-1/2}\right)\text{ uniformly in }d.$$

In particular, we show convergence of the *fidi* and tightness. For the former, the numerator of $T^{1/2-d^0}\left(\hat{\mu}\left(d^0\right)-\mu^0\right)$ can be written for any $d\neq d^0$ as

$$T^{-1/2-d^0+2d}\sum_{k=1}^{T}\left(\sum_{j=0}^{T-k}\pi_j\left(d-d^0\right)\pi_{k+j}\left(d-1\right)\right)u_k,$$

with a variance of

$$T^{-1-2d^0+4d}\sum_{k=1}^{T}\left(\sum_{j=0}^{T-k}\pi_j\left(d-d^0\right)\pi_{k+j}\left(d-1\right)\right)^2 \tag{1.40}$$

which is of order $O\left(1\right)$ and the denominator

$$T^{2d-1}\sum_{t=1}^{T}\pi_{t-1}^2\left(d-1\right)\rightarrow\frac{1}{\Gamma^2\left(1-d\right)\left(1-2d\right)}. \tag{1.41}$$

For tightness we show that

$$E|T^{1/2-d^0}\left(\hat{\mu}\left(d_2\right)-\mu^0\right)-T^{1/2-d^0}\left(\hat{\mu}\left(d_1\right)-\mu^0\right)|^2\leq K|d_2-d_1|^2. \tag{1.42}$$

From a Taylor approximation we find that the term in the bracket of LHS of (1.42)

$$\leq |d_2-d_1|T^{1/2-d^0}\frac{\partial}{\partial d}\left(\hat{\mu}\left(d^*\right)-\mu^0\right).$$

Therefore, it suffices to show that

$$E|\left[T^{1/2-d^0}\frac{\partial}{\partial d}\left(\hat{\mu}\left(d^*\right)-\mu^0\right)\right]^2|=C<\infty.$$

We can express the term in brackets as

$$\begin{aligned}
& T^{1/2-d^0}\frac{\partial}{\partial d}\left(\hat{\mu}\left(d^*\right)-\mu^0\right) \\
= \; & T^{1/2-d^0}\frac{\partial}{\partial d}\frac{\sum_{k=1}^{T}\left[\sum_{j=0}^{T-k}\pi_j(d^*-d^0)\pi_{k+j}(d^*-1)\right]u_k}{\sum_{t=1}^{T}\pi_{t-1}^2(d^*-1)} \\
= \; & T^{1/2-d^0}\frac{\sum_{k=1}^{T}\left[\sum_{j=0}^{T-k}\dot{\pi}_j(d^*-d^0)\pi_{k+j}(d^*-1)\right]+\left[\sum_{j=0}^{T-k}\pi_j(d^*-d^0)\dot{\pi}_{k+j}(d^*-1)\right]u_k}{\sum_{t=1}^{T}\pi_{t-1}^2(d^*-1)} \qquad (1.43) \\
& -T^{1/2-d^0}\frac{2\sum_{t=1}^{T}\pi_{t-1}(d^*-1)\dot{\pi}_{t-1}(d^*-1)\sum_{k=1}^{T}\left[\sum_{j=0}^{T-k}\pi_j(d^*-d^0)\pi_{k+j}(d^*-1)\right]u_k}{\left[\sum_{t=1}^{T}\pi_{t-1}^2(d^*-1)\right]^2} \qquad (1.44)
\end{aligned}$$

First, when taking the expectation of the square of the second term (1.44) we find

$$\begin{aligned}
& E\left[T^{1/2-d^0}\frac{2\sum_{t=1}^{T}\pi_{t-1}(d^*-1)\dot{\pi}_{t-1}(d^*-1)\sum_{k=1}^{T}\left[\sum_{j=0}^{T-k}\pi_j(d^*-d^0)\pi_{k+j}(d^*-1)\right]u_k}{\left[\sum_{t=1}^{T}\pi_{t-1}^2(d^*-1)\right]^2}\right]^2 \\
= \; & T^{1-2d_0}\frac{\left(\sum_{t=1}^{T}\pi_{t-1}(d^*-1)\dot{\pi}_{t-1}(d^*-1)\right)^2}{\left[\sum_{t=1}^{T}\pi_{t-1}^2(d^*-1)\right]^2}\frac{\sum_{k=1}^{T}\left[\sum_{j=0}^{T-k}\pi_j(d^*-d^0)\pi_{k+j}(d^*-1)\right]^2}{\left[\sum_{t=1}^{T}\pi_{t-1}^2(d^*-1)\right]^2}.
\end{aligned}$$

From (1.40) and (1.41), the second factor is of order T^{2d^0-1}. The denominator of the first factor is of order $T^{2(1-2d^*)}$. Defining similarly as in Lasak (2009),

$$\pi_j^*\left(d^*-d^0\right)=T^{d^*-d^0}\pi_j\left(d^*-d^0\right),$$

with

$$T^{d^0-d^*}|\dot{\pi}_j^*\left(d^*-d^0\right)|\leq Kj^{d^0-d^*-1}\ln\left(\frac{j}{T}\right), \tag{1.45}$$

which is a consequence of equation (19) of Lasak (2009) and using that $\int_0^1 x^{-2d}\ln x dx<\infty$ for $d<1/2$, we find that the numerator of the first factor is of order $O\left(T^{1-2d^*}\right)$ since

$$\begin{aligned} T\frac{1}{T}\sum_{t=1}^{T}\pi_{t-1}(d^*-1)\dot{\pi}_{t-1}(d^*-1) &\leq KT^{1-2d^*}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t-1}{T}\right)^{-d^*}\left(\frac{t-1}{T}\right)^{-d^*}\ln\left(t/T\right) \\ &= O\left(T^{1-2d^*}\right) \end{aligned}$$

Hence, the expectation of the square of the second term (1.44) is bounded. Next, when taking expectation of the square of term (1.43), we find

$$\begin{aligned} &E\left[T^{1/2-d^0}\frac{\sum_{k=1}^{T}\left[\sum_{j=0}^{T-k}\dot{\pi}_j(d^*-d^0)\pi_{k+j}(d^*-1)\right]+\left[\sum_{j=0}^{T-k}\pi_j(d^*-d^0)\dot{\pi}_{k+j}(d^*-1)\right]u_k}{\sum_{t=1}^{T}\pi_{t-1}^2(d^*-1)}\right]^2 \\ =\;& T^{1-2d^0}\frac{\sum_{k=1}^{T}\left(\sum_{j=0}^{T-k}\dot{\pi}_j(d^*-d^0)\pi_{k+j}(d^*-1)+\pi_j(d^*-d^0)\dot{\pi}_{k+j}(d^*-1)\right)^2}{\left[\sum_{t=1}^{T}\pi_{t-1}^2(d^*-1)\right]^2}. \end{aligned} \tag{1.46}$$

When applying (1.45) with argument d^*-d^0 to the numerator of expression (1.46),

$$\begin{aligned} & T^{3+2d^0-2d-2-2d}\frac{1}{T}\sum_{k=1}^{T}\left(\frac{1}{T}\sum_{j=0}^{T-k}K_1\frac{j}{T}^{d^0-d-1}\ln\left(j/T\right)\left(\frac{k+j}{T}\right)^{-d}\right. \\ & \left.+K_2\frac{j}{T}^{d^0-d-1}\ln\left(j/T\right)\left(\frac{k+j}{T}\right)^{-d}\right)^2 \\ \simeq\;& T^{1+2d^0-4d}\int_0^1\left(\int_0^{1-K}J^{d^0-d-1}\ln J\left(K+J\right)^{-d}\right)^2=O_p\left(T^{-1+2d^0-4d^*}\right). \end{aligned}$$

Combining this result with the order $O\left(T^{2(1-2d^*)}\right)$ of the denominator of expression (1.46), we confirm that the expectation of the square of the first term is bounded as well. Since the cross term is also bounded, we obtain the result.

Proof of Theorem 3

This proof follows closely the proof of Theorem 2 of Boldea and Hall (2010). We consider the case of *three* breaks.We analyze three different cases of changing parameters that require a different analysis:

- case A: break in memory and mean, or break in memory
- case B: break in mean; $d_1^0 = d_2^0 = d_3^0 \geq 0$

Consistency of the three breaks is already established. Because of consistency we only have to consider the behavior of the break points in

$$V_{\in} = \left\{ (T_1, T_2, T_3) : |T_i - T_i^0| \leq \varepsilon T \ (i = 1, 2, 3) \right\}.$$

First, consider case $\hat{T}_2 < T_2^0$. In contrast to Boldea and Hall (2009), here the argument is not symmetric and we have to consider also the case $T_2^0 > T_2$. The proof works basically by showing that the break point is with a very small probability in the set

$$V_{\in}(C) = \left\{ (T_1, T_2, T_3) : |T_i - T_i^0| \leq \varepsilon T \ (i = 1, 2, 3); \ \Delta_2 = T_2^0 - T_2 > C \right\}.$$

Hence with large probability $|\hat{T}_2 - T_2^0| < C$. We will show that if $T_2 \in V_{\in}(C)$,

$$P \left\{ \min_{V_{\in}(C)} \frac{S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)}{\Delta_2^{\delta}} \leq 0 \right\} < \eta, \text{ for } T \geq T(\eta) \tag{1.47}$$

contradicting the sum of squares minimization and implying that T_2 does not belong to $V_{\in}(C)$. For case A, $\delta = 1$ and, for case B, $\delta = 1 - 2d_2^0$. We show that

$$S_T(T_1, T_2, T_3) - S_T\left(T_1, T_2^0, T_3\right) = (SSR1 - SSR3) - (SSR2 - SSR3)$$

is positive with high probability for large T picking ε and C where

$$\begin{aligned} SSR1 &= S_T(T_1, T_2, T_3), SSR2 = S_T\left(T_1, T_2^0, T_3\right) \text{ and} \\ SSR3 &= S_T\left(T_1, T_2, T_2^0, T_3\right). \end{aligned}$$

Lemma 4 discusses the rates of convergence of the estimators $(\theta_1^*, \theta_2^*, \theta_3^{**}, \theta_4^*), (\theta_1^*, \theta_2^{**}, \theta_3^*, \theta_4^*)$ and $(\theta_1^*, \theta_2^{***}, \theta_2^{\delta}, \theta_3^{***}, \theta_4^*)$ for the first, second and third partition respectively.

We locate the dominating terms in $S_T(T_1, T_2, T_3) - S_T(T_1, T_2^0, T_3)$ and show that at least some are positive with large probability. Equation (1.47) is equivalent to

$$\Delta_2^{-\delta}(SSR1 - SSR2) \sim O_p^+(1) \tag{1.48}$$

Next, we introduce some notation:

$$I_1 = [1, T_1]; I_2 = [T_1 + 1, T_2], I_2^{\Delta} = [T_2 + 1, T_2^0], \ I_3 = [T_2^0 + 1, T_3], \ I_4 = [T_3 + 1, T]$$

and we get to

$$\begin{aligned}\frac{SSR_1 - SSR_3}{\Delta_2^\delta} &= \frac{1}{\Delta_2^\delta}\left[\sum\nolimits_{I_2^\Delta}\left[u_t^2\left(\theta_3^{**}\right) - u_t^2\left(\theta_2^\delta\right)\right] + \sum\nolimits_{I_3}\left[u_t^2\left(\theta_3^{**}\right) - u_t^2\left(\theta_3^{***}\right)\right]\right] \\ &= D_1 + D_2.\end{aligned}$$

Since θ_3^{**} estimates θ_3^0 and θ_2^δ estimates θ_2^0, there is a mismatch in D_1, while there is none in D_2 (θ_3^{**} and θ_3^* estimate θ_3^0). D_1 should therefore in the limit dominate D_2. In particular,

$$\begin{aligned}D_1 &= \frac{1}{\Delta_2^\delta}\sum\nolimits_{I_2^\Delta}\left[u_t^2\left(\theta_3^{**}, d_2^0\right) - u_t^2\left(\theta_2^\delta, d_2^0\right)\right] = \frac{1}{\Delta_2^\delta}\sum\nolimits_{I_1^\Delta} d_t^2\left(\theta_3^{**}, d_2^0\right) - \frac{1}{\Delta_2^\delta}\sum\nolimits_{I_2^\Delta} d_t^2\left(\theta_2^\delta, d_2^0\right) \\ &\quad -\frac{1}{\Delta_2^\delta}\sum\nolimits_{I_2^\Delta} d_t\left(\theta_3^{**}, d_2^0\right)u_t + \frac{1}{\Delta_2^\delta}\sum\nolimits_{I_2^\Delta} d_t\left(\theta_2^\delta, d_2^0\right)u_t \\ &= D_{1,1} - D_{1,2} - D_{1,3} + D_{1,4}\end{aligned}$$

We show that

$$D_1 = \frac{1}{\Delta_2^\delta}\sum\nolimits_{I_2^\Delta}\left[u_t^2\left(\theta_3^{**}\right) - u_t^2\left(\theta_2^\delta\right)\right]$$

has a limit in probability of order $O_p^+(1)$. From Lemma 5 (any θ) and some modification of it, these ULLN hold at the appropriate rate. It remains to show that the limits behave like claimed. In particular, for the term $D_{1,1}$, we find for case A,

$$D_{1,1} = \frac{1}{\Delta_2}\sum\nolimits_{I_2^\Delta} d_t^2\left(\theta_3^{**}, \theta_2^0\right) \xrightarrow{p} \sum_{j=1}^{\infty}\pi_j^2(d_3^{**} - d_2^0) > 0$$

and for case B from Lemma 6 and some modification of it,

$$D_{1,1} = \frac{1}{\Delta_2^{1-2d_2^0}}\sum\nolimits_{I_2^\Delta} d_t^2\left(\theta_3^{**}, \theta_2^0\right) \xrightarrow{p} \frac{\left(\mu_2^0 - \mu_3^{**}\right)^2}{\Gamma^2\left(1-d^0\right)\left(1-2d^0\right)} > 0$$

where convergence follows from Lemma 1 respectively and the last inequality follows from $d_3^{**} \xrightarrow{p} d_3^0$ and $|d_2^0 - d_3^0| > \varepsilon$. Further, for case A,

$$D_{1,2} = \frac{1}{\Delta_2}\sum\nolimits_{I_2^\Delta} d_t^2\left(\theta_2^\partial, \theta_2^0\right) \xrightarrow{p} \sum_{j=1}^{\infty}\pi_j^2(d_2^\partial - d_2^0) = o_p(1)$$

and for case B,

$$D_{1,2} \simeq \frac{1}{\Delta_2^{1-2d_2^0}}\left(\mu_2^0 - \mu_2^\delta\right)^2\frac{\Delta_2^{1-2d_2}}{\Gamma^2\left(1-d_2\right)\left(1-2d_2\right)} \xrightarrow{p} \frac{\left(\mu_2^0 - \mu_2^\delta\right)^2}{\Gamma^2\left(1-d^0\right)\left(1-2d^0\right)} = o_p(1)$$

for $\delta = 1$ and $\delta = 1 - 2d^0$ where the convergence follows again from Lemma 5 and 6 respectively and the last step follows from $\left|d_2^\partial - d_2^0\right| = O_p(\Delta_2^{-1/2})$ and for C sufficiently large. Finally, we have again from Lemma 5 and 6 uniform convergence of infinite sums of $\{u_t d_t\}$ implying that

$$\frac{1}{\Delta_2^{\delta}} \sum\nolimits_{I_2^{\Delta}} d_t \left(\theta, \theta_2^0\right) u_t = o\left(1\right)$$

implying that $D_{1,3} = o\left(1\right)$ and $D_{1,4} = o\left(1\right)$. Combining the four terms we obtain $D_1 \sim O_p^+\left(1\right)$.

As in Boldea and Hall (2009), the term D_2 is different since we sum over a different interval: I_3 with an an interval length of order T instead of I_2^{Δ}. Hence, this will be the critical term for getting the appropriate rate of convergence of the break fraction. In particular, as consequence of Lemma 5 in the case A and Lemma 6 the term is of order $o_p\left(1\right)$.

Finally, as in Boldea and Hall (2009), we can show that the term SSR2-SSR3 are of order $o_p\left(1\right)$.

Chapter 2

Lagrange Multiplier and Wald Tests for a Change in the Persistence and Level of a Time Series

(joint with Juan J. Dolado and Carlos Velasco)

Abstract. This paper analyzes Lagrange Multiplier (LM) and Wald tests for breaks in the memory and the level of a time series. Its contribution is to consider both types of breaks simultaneously to solve a potential confounding problem between long memory changes and breaks in the level and in persistence. Identifying different sources of breaks is relevant for several reasons, such as improved forecasting, shock identification and avoiding spurious fractional cointegration. We derive the asymptotic distribution for both known and unknown break fraction as well as for known and unknown parameters under the null and a local break hypothesis. Further, we extend the proposed testing procedures by allowing for potentially breaking short-run dynamics.

2.1 Introduction

Since the work by Granger and Hyung (2004) there has been a long discussion on whether some time series are truly driven by fractionally integrated processes, FI(d), or their long memory (d) is spuriously generated by breaks in their level (see e.g., Lobato and Savin, 1998, Mikosch and Starica, 2004, and Perron and Qu, 2010). Furthermore, stochastic processes with breaks in the memory parameter could also generate time series which resemble financial time series found in practice (see McCloskey, 2010). The debate has given rise to two strands in this literature. The first one has focused on testing directly for breaks in the degree of fractional integration. Motivated by the popular rationalization of FI(d) processes provided by Granger and Joyeux (1978), it has been argued that changes in monetary policy regimes or in financial regulations, sectorial composition, etc., may have led to shifts in the distribution of the persistence parameters underlying the disaggregated components of many macro and financial variables (inflation, unemployment, GDP, squared financial returns, etc.). Thus, this could have altered the long memory properties of the aggregates over different subsamples (see, eg. Gadea and Mayoral, 2005). Accordingly, several tests for the null of a constant value of d against the alternative of breaks at known or unknown dates have been developed in semiparametric and parametric setups (see, e.g., Beran and Terrin, 1996, Hassler and Meller, 2011, Hassler and Scheithauer, 2011, Martins and Rodrigues, 2010, McCloskey, 2010, and Sibbertsen and Kruse, 2009).[1] The second strand centers on testing for a break in the level with a stationary long-memory error term, yet without allowing for breaks in d (see e.g., Hsu and Kuan, 1998**,** Lavielle and Moulines, 2000, and Shao, 2011).

However, to our knowledge, not much has been done about the possibility of considering joint breaks in d and the level and, in such a case, disentangling whether the change occurs either in one of two parameters or in both. Hassler and Meller (2011) extend Breitung and Hassler's (2002) LM test to deal with breaks in d allowing for level shifts but this is done in a sequential way. First, the location of the mean break is detected using Hsu's (2005) semiparametric testing procedure and next the original time series is demeaned (with a broken intercept) to test for a break in d. How the two-step procedure affects the asymptotic properties of the test on d is not formally proved. For example, when demeaning, the mean would be very imprecisely estimated when d is close to 0.5 because of its $T^{1/2-d}$ rate of convergence. Therefore, if results depend on correct demeaning, there is an additional uncertainty which is not taken into account in the semiparametric literature.

Indeed, as far as we know, the only paper that addresses the issue of allowing for

[1] Forerunners of these tests are the approaches proposed by Kim, Belaire-Franch and Amador (2002) and Busetti and Taylor (2004) to test for I(0) series against alternatives where there is a change from I(0) to I(1).

a joint break is Chapter 1 where I propose a unified approach for modeling breaks in level and memory by extending Bai and Perron´s (1998) methodology, based on supF tests with unknown break fraction, to an FI(d) context. He also analyzes least squares estimation of (multiple) breaks in this setup. Under $d \in [0, 0.5)$, he discusses a linear model with multiple breaks. Consistency and $T-$rate convergence of the break-fraction estimator and the asymptotic distribution of the parameter estimates in the different regimes are provided under different magnitudes of breaks. Finally, he also provides a series of supF test statistics for the existence and number of breaks.

As discussed in Dolado et al. (2005), it is important to distinguish between long memory, breaks in the memory parameter and breaks in the level for several reasons. First, there is the improved forecasting. In particular, the higher the memory is, the more observations need to be used to produce good forecasts. Further, forecasting requires some knowledge on the stability of the series. Second, there is the identification of shocks. For economic modelling it matters whether the underlying shocks are persistent or transitory. Take e.g. the inflation rate as a measure of the credibility of the central bank. The less persistent the shocks are, the more credible is the central bank. Finally, in order to model two series as fractionally cointegrated, both series have to have the same memory. Thus, if the memory is estimated too high due to instabilities, a discovered fractional cointegration could be spurious.

In order to further motivate the use of a FI(d) model and the joint modelling of breaks in level and in memory in particular, consider the following example regarding the time-series behavior of the inflation rate. Let us suppose that individual firms $(i = 1, 2, .., N)$ set prices in period t, p_t^i according to the well-known Calvo´s (1983) model,

$$p_t^i = \theta^i p_{t-1}^i + (1 - \theta^i) p_t^{i^*}, i = 1, .., N,$$

where firm i is allowed to choose its optimal price level $p_t^{i^*}$ with probability $(1 - \theta^i)$, while it keeps the price set in the previous period with probability θ^i. For simplicity, let us assume that the optimal price evolves as a random walk with drift,

$$\Delta p_t^{i^*} = \mu^i + \xi_t^i.$$

Then, combining the following two expressions,[2]

$$\Delta p_t^i = \left(1 - \theta^i\right) (p_t^{i^*} - p_{t-1}^i),$$

[2]The first expression is derived from subtracting the identity $p_t^i \equiv \theta^i p_{t-1}^i + (1 - \theta^i) p_{t-1}^i$ from the above-mentioned price setting equation.

$$p_t^i - p_t^{i^*} = -\theta^i \Delta p_t^{i^*} + \theta^i (p_{t-1}^i - p_{t-1}^{i^*}),$$

yields

$$\Delta p_t^i = (1-\theta^i)\mu^i + \theta^i \Delta p_{t-1}^i + (1-\theta^i)\xi_t^i.$$

Next, consider aggregate inflation to be defined as

$$\Delta p_t = \Sigma_i \Delta p_t^i,$$

with $\theta^i \sim \mathcal{F}(\theta, d) \sim c\theta^{-d}$ where c$\in (0, \infty)$ as $\theta \uparrow 0^+$. Then, as shown by Zaffaroni (2004), under these assumptions $\Delta p_t \sim$FI(d). Now, a change in the parameter θ^i, due e.g. to more or less competition in the product and services markets, implies changes in persistence and level, while a change in the parameter μ^i, due e.g. to a change in monetary policy, implies a sole change in the level. Finally, a change in θ^i and μ^i in an offsetting manner, i.e., such that $(1-\theta^i)\mu^i$ remains invariant, leads to a change in the persistence only. Thus, this simple example illustrates the different types of parameter shifts that we analyze here.

In this paper, we contribute to this sparse literature by deriving the corresponding LM and Wald tests for the joint hypothesis of constant d and level using a parametric setup. As discussed extensively in the literature, semiparametric estimation of d would help us to abstract from short term dynamics while concentrating on the memory estimation. Yet, parametric estimation has a higher rate of convergence assuming the model is correctly specified and, what is more relevant, it provides the opportunity to include potential breaks as well in the short term dynamics, and to quantify the effect in the power of the tests coming from the estimation of memory and level under the null. LM tests are computationally attractive because they require only estimation under the null, while Wald tests can exploit further information on the alternative, potentially leading to higher power.

The rest of Chapter 2 is structured as follows. In Section 2, we lay out the data generating processes (DGP) and derive the asymptotic properties of the LM and Wald tests for a single break both under the null and under local and fixed alternatives, distinguishing among different setups: known and unknown breaking date, change in only one of the parameters, etc. In Section 3, we compare the behavior of the LM, Wald and an alternative likelihood ratio test under the fixed alternative and show that there is an asymptotic inequality between them. In Section 4, we provide generalizations of the proposed tests to detect multiple breaks and discuss bootstrap procedures to improve their finite sample performance and conclude. The proofs are collected in the Appendix.

2.2 LM and Wald Tests

For simplicity, we start focusing on the case of *a single* break. It can be readily generalized to more breaks along the lines of Chapter 1 as discussed in Section 2.4. In particular, we assume that the time series is FI(d_0) with $d_0 \in D_1$ or D_2, where $D_1 = (-0.5, 0]$ and $D_2 = [0, 0.5)$ for $t = 1, ..., [\lambda_0 T]$ and FI(d_1) with $d_1 \in D_1$ or D_2 for $t = [\lambda_0 T] + 1, ..., T$ with the level of the series being μ_0 in the former subsample and μ_1 in the latter. The parameter λ_0 denotes the true break fraction and lies in the interval $(\epsilon, 1 - \epsilon)$, where $\epsilon > 0$. In particular, we consider the following smooth transition

$$\begin{aligned} \Delta_t^{d_0} (y_t - \mu_0) = \varepsilon_t,\ t = 1, ..., [\lambda_0 T] \\ \Delta_t^{d_1} (y_t - \mu_1) = \varepsilon_t,\ t = [\lambda_0 T] + 1, ..., T, \end{aligned} \tag{2.1}$$

where $\Delta_t^{-b} = \sum_{k=0}^{t-1} \pi_k(-b) L^k$, with $\pi_k(-b) = \frac{\Gamma(k+b)}{\Gamma(b)\Gamma(k+1)}, k = 0, ..., t-1$, denotes the (truncated or Type-II) fractional-differencing filter. Notice that, unlike the DGP discussed in Chapter 1 in a similar setup to the one here, here both memory and mean are gradually changing. Given the persistent nature of the process, depending on the size of the memory parameters, the transition of the mean may occur very slowly. In particular, the data is generated recursively as

$$y_t = \begin{cases} \mu_0 + (1 - \Delta_t^{d_0})(y_t - \mu_0) + \varepsilon_t = \mu_0 + \Delta_t^{-d_0} \varepsilon_t,\ t = 1, ..., [\lambda_0 T] \\ \mu_1 + (1 - \Delta_t^{d_1})(y_t - \mu_1) + \varepsilon_t,\ t = [\lambda_0 T] + 1, ..., T. \end{cases}$$

Also notice that, for a nonstationary process with a memory $d_0, d_1 \in (1/2, 1]$ or $[1, 3/2)$, and a potentially breaking linear trend,

$$\Delta_t^{d_i} (y_t - \mu_i - \beta_i t) = \varepsilon_t,\ i = 0, 1,$$

our method can be applied to the first-differenced data to test for breaks in β.

The following notation will be used in the sequel. As regards the memory parameter, d_0 denotes the true memory if not changing and likewise for the level parameter μ_0. In turn, d_1 and μ_1 denote the true memory and level parameters in the second subsample, respectively, when they shift during the latter. Finally, θ and ν denote a shift from d_0 and μ_0, respectively. Consider the following model

$$\begin{aligned} \Delta_t^{d_0} (y_t - \mu_0) = \varepsilon_t,\ t = 1, ..., [\lambda_0 T] \\ \Delta_t^{d_0+\theta} (y_t - \mu_0 - \nu) = \varepsilon_t,\ t = [\lambda_0 T] + 1, ..., T, \end{aligned} \tag{2.2}$$

where δ and η (normalized by $T^{-0.5}$ and $T^{0.5-d_0}$) are the corresponding shifts under local alternatives.

2.2.1 LM Tests

In particular, according to the LM testing principle, we test for the null

$$H_0 : \theta = \nu = 0 \tag{H0}$$

against the alternative that, at an (unknown) fraction λ_0 of the sample size,

$$H_1(\lambda_0) : \theta \neq 0 \text{ and/or } \nu \neq 0. \tag{H1}$$

Only to derive the likelihood, it is assumed that $\varepsilon_t \sim NID(0, \sigma^2)$. With $D_t(\lambda_0) = 1(t > \lambda_0 T)$, the likelihood function can then be written as,

$$L\left(\theta, d, \nu, \mu, \sigma^2, \lambda\right) = -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left\{\Delta_t^{d+\theta D_t(\lambda)}(y_t - \mu) - D_t(\lambda)\Delta_t^{d+\theta D_t(\lambda)}\nu\right\}^2. \tag{2.3}$$

Given a break fraction λ, the LM test is based on the derivatives of $L(\theta, d, \nu, \mu, \sigma^2, \lambda)$ in direction of θ and ν, evaluated at the restricted estimates $(0, \tilde{d}_T, 0, \tilde{\mu}_T, \tilde{\sigma}_T^2, \lambda)$, and reads as follows,

$$LM^{\vartheta}(\lambda) = \frac{\partial L}{\partial \psi}|_{\theta=0,\nu=0}\left(-\frac{\partial^2 L}{\partial \psi^2}|_{\theta=0,\nu=0}\right)^{-1}\frac{\partial L}{\partial \psi'}|_{\theta=0,\nu=0}, \tag{2.4}$$

where the first and second derivatives of the likelihood function are evaluated under H_0 and where $\vartheta \in \{d; \mu; (d, \mu)\}$ denotes the set of parameters in which we test for breaks. In particular,

$$\frac{\partial L}{\partial \psi}|_{\theta=0,\nu=0} = \begin{pmatrix} T\sum_{k=1}^{T-1}\frac{1}{k}\tilde{\rho}_k^*(\tilde{\varepsilon}_t) \\ \frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}(\Delta_t^{\tilde{d}_T}1)\tilde{\varepsilon}_t \end{pmatrix},$$

where

$$\tilde{\rho}_k^*(\tilde{\varepsilon}_t) = \frac{\tilde{\sigma}_T^{-2}}{T}\sum_{t=\max\{[\lambda T]+1-k,1\}}^{T-k}\tilde{\varepsilon}_t\tilde{\varepsilon}_{t+k},$$

such that,

$$\tilde{\varepsilon}_t = \Delta_t^{\tilde{d}}(y_t - \tilde{\mu}_T)1\{t > 0\}, \tag{2.5}$$

and the variance estimator is,

$$\tilde{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T} \tilde{\varepsilon}_t^2.$$

Next, the Hessian becomes,

$$\frac{\partial^2 L}{\partial \psi^2}|_{\theta=0,\nu=0} = \begin{pmatrix} \lambda \frac{1}{\tilde{\sigma}_T^2} \sum\limits_{t=[\lambda T]+1}^{T} (\log \Delta \tilde{\varepsilon}_t)^2 & o_p(T^{1-\tilde{d}_T}) \\ o_p(T^{1-\tilde{d}_T}) & \lambda^{1-2\tilde{d}_T} \frac{1}{\tilde{\sigma}_T^2} \sum\limits_{t=[\lambda T]+1}^{T} (\Delta_t^{\tilde{d}_T} 1)^2 \end{pmatrix}.$$

Given a break fraction λ, we estimate the level μ and the memory d under the constraints of $\theta = \nu = 0$. If the break fraction is considered to be unknown, then the break fraction is determined by

$$\hat{\lambda} = \arg\max_{\lambda} LM^{\vartheta}(\lambda). \tag{2.6}$$

We derive the asymptotic distributions of the tests under the following assumptions:

Assumption 1 $\varepsilon_t \sim iid\,(0, \sigma^2)$ with $q > max\{2, \frac{2}{1-2d_0}\}$ moments.

Assumption 2a The memory parameter $\mathrm{d}_0 \in (-1/2, 0]$.

Assumption 2b The memory parameter $\mathrm{d}_0 \in [0, 1/2)$.

To assess the local power, we study the properties of the LM under a local break alternative,

$$H_{1T}(\lambda_0) : \theta_0 = \delta/T^{1/2}, \nu_0 = \eta/T^{1/2-d_0}, \tag{2.7}$$

where notice that the critical rate in the direction of the level differs from the standard $\sqrt{T}$ rate.

Theorem 22 discusses the asymptotic distribution of the test (2.4), when the break fraction is considered to be unknown and has to be estimated along with the other parameters.

Theorem 22 *Under Assumptions 1 and 2a or 2b and under the local hypothesis*

$H_{1T}(\lambda_0)$ and an unknown break fraction λ_0,

$$
\begin{aligned}
\sup_{\lambda} LM_T(\lambda) \xrightarrow{d} \sup_{\lambda} & \left\{ \left[\frac{\left(\lambda B(1) - B(\lambda) - \delta\left(\lambda(1-\lambda_0) - (\lambda-\lambda_0)_+\right)\frac{\pi}{\sqrt{6}}\right)^2}{\lambda(1-\lambda)} \right] \right. \\
& \left. + \left[\frac{\left(\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}(1) - \tilde{W}_{1/2-d_0}(\lambda) - \eta \frac{\lambda^{1-2d_0}\left(1-\lambda_0^{1-2d_0}\right) - \left(\lambda^{1-2d_0} - \lambda_0^{1-2d_0}\right)_+}{\sqrt{1-2d_0}\Gamma(1-d_0)}\right)^2}{\lambda^{1-2d_0}\left(1-\lambda^{1-2d_0}\right)} \right] \right\} \\
\equiv \sup_{\lambda} & \left\{ [I^d(\lambda)] + [II^{\mu}(\lambda)] \right\},
\end{aligned}
$$

where $(f)_+$ denotes the positive part of f and where $\tilde{W}_{1/2-d_0}(\lambda) = \int_0^{\lambda} s^{-d_0} dB'(s)$ is a variant of fractional Brownian Motion and $B(s)$ and $B\prime(s)$ are two independent Brownian motions.

The limiting distribution derived in Theorem 1 is a function of both standard Brownian Motion (BM) and a variant of fractional BM (fBM). Notice that $\tilde{W}_{1/2-d_0}(\lambda)$ differs from the standard fBM $W_{1/2-d_0}(\lambda) = \int_0^{\lambda}(\lambda - s)\, ds$ by its particular covariance structure, given by,

$$
Cov\left(\tilde{W}_{1/2-d_0}(s), \tilde{W}_{1/2-d_0}(t)\right) = \frac{\min(s,t)^{1-2d_0}}{\Gamma(1-d_0)(1-2d_0)}.
$$

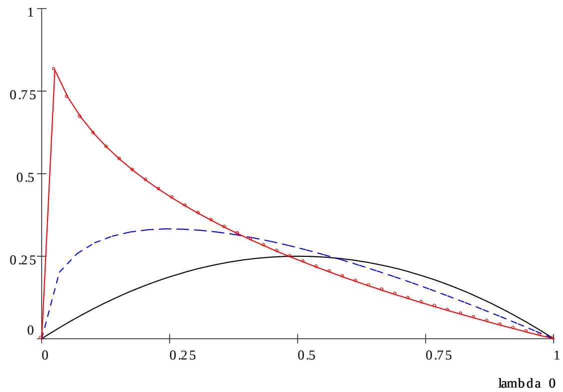
Remark 1. *The two components $[I^d]$ and $[II^{\mu}]$ in the asymptotic distribution of the sup-LM test capture the contributions of the limiting distributions of the local shift and its unknown breaking time of the memory parameter $[I^d]$ and of the level and its unknown time of break $[II^{\mu}]$, respectively.*

It can be noticed that while $[I^d]$ is symmetric around the break fraction $\lambda_0 = 1/2$, $[II^{\mu}]$ is positive skewed. Hence, due to this feature, the local power is highest for $\lambda_0 = 1/2$, if we consider only a break in the level, but is highest for some $\lambda_0 < 1/2$, if we consider breaks in memory or in both memory and level. Figure 2-1 depicts the local power for a break in the level as a function of the break fraction λ_0 for $\eta = 1$ and for $d = 0, 0.25, 0.45$.

Theorem 1 also nests the special cases of testing for a break only in the memory (where $[II^{\mu}]$ disappears) or only in the level (where $[I^d]$ vanishes). Corollary 1, in turn, provides the asymptotic distribution for the LM test when the break fraction λ is assumed to be known.

Corollary 23 *Under Assumptions 1 and 2a or 2b and under the local hypothesis*

Figure 2-1: **Tests for breaks in the level: Local power in function of the break fraction**



$\eta = 1$, $d_0 = 0$ (black solid), $d_0 = 0.25$ (blue dashed) and $d_0 = 0.45$ (red dotted)

$H_{1T}(\lambda_0)$, *a known break fraction* λ_0,

$$LM_T \xrightarrow{d} \chi_2^2(c),$$

with non-centrality parameter

$$c = \delta^2 \lambda_0 (1 - \lambda_0) \frac{\pi^2}{6} + \eta^2 \frac{\lambda_0^{1-2d_0}(1 - \lambda_0^{1-2d_0})}{(1 - 2d_0)\, \Gamma^2 (1 - d_0)}.$$

As expected, in this case, the asymptotic distribution is a $\chi_2^2(c)$ with the non-centrality parameter c under local alternatives, which is a function of the two drifts. As before, Corollary 23 nests the cases of testing for a break only in d or only the level. The limit distribution becomes $\chi_1^2(c)$, where the second term in the definition of c drops for the former even if $\eta \neq 0$ and the second term does so for the latter even if $\delta \neq 0$.

Proposition 1 illustrates the intuitive result that, when the memory and the level are assumed to be known rather than estimated, the local power increases because the variance decreases. The cases of known μ_0 and unknown d_0 and vice versa follow from combining Proposition 24 with Theorem 22 and with Corollary 23 respectively.

Proposition 24 *Known parameters.*
a) Under Assumptions 1 and 2a or 2b and under the local hypothesis $H_{1T}(\lambda_0)$*, an*

unknown break fraction λ_0, and a priori knowledge of d_0 and μ_0,

$$\sup_{\lambda} LM_T(\lambda) \xrightarrow{d} \sup_{\lambda} \left\{ \frac{\left(B(1) - B(\lambda) - \delta\left(1 - \max\{\lambda, \lambda_0\}\right) \frac{\pi}{\sqrt{6}} \right)^2}{(1-\lambda)} \right.$$
$$\left. + \frac{\left(\tilde{W}_{1/2-d_0}(1) - \tilde{W}_{1/2-d_0}(\lambda) - \eta \frac{1 - \max\{\lambda, \lambda_0\}^{1-2d_0}}{\sqrt{1-2d_0}\Gamma(1-d_0)} \right)^2}{\left(1 - \lambda^{1-2d_0}\right)} \right\}.$$

b) Under Assumptions 1 and 2a or 2b and under the local hypothesis H_{1T}, a known break fraction λ_0, and a priori knowledge of d_0 and μ_0,

$$LM_T \xrightarrow{d} \chi_2^2(c),$$

with noncentrality parameter

$$c = \delta^2 (1 - \lambda_0) \frac{\pi^2}{6} + \eta^2 \frac{1 - \lambda_0^{1-2d_0}}{(1 - 2d_0)\Gamma^2(1 - d_0)}.$$

Thus, if the parameters in the first regime are known, the local drift increases and the null distribution changes as well.

Stable short run dynamics

Next, Proposition 25 provides the asymptotic behavior of the LM test when there are short run dynamics in form of a stable autoregressive structure of known lag length. In particular, assume that the DGP is,

$$\begin{gathered} \alpha(L) \Delta_t^{d^0} (y_t - \mu^0) = \varepsilon_t,\ t = 1, ..., [\lambda_0 T] \\ \alpha(L) \Delta_t^{d^0+\theta} (y_t - \mu^0 - \nu) = \varepsilon_t,\ t = [\lambda_0 T] + 1, ..., T, \end{gathered} \tag{2.8}$$

where $\alpha(L) = 1 - \alpha_1 L - \cdots \alpha_p L^p$ is a stable polynomial of order p with unknown coefficients. Let $\omega^2 = \frac{\pi^2}{6} - \kappa' \Phi^{-1} \kappa$, such that $\kappa = (\kappa_1, ..., \kappa_p)'$ with $\kappa_k = \sum_{j=k}^{\infty} j^{-1} c_{j-k}$, $k = 1, ..., p$, where the c_j are the coefficients of L^j in the expansion of $1/\alpha(L)$ and where $\Phi = [\Phi_{k,j}]$, $\Phi_{k,j} = \sum_{t=0}^{\infty} c_t c_{t+|k-j|, k,j=1,...,p}$ denotes the Fisher information matrix for α under Gaussianity. From the estimation under the null, we obtain the residuals

$$\tilde{\varepsilon}_t = \tilde{\alpha}(L) \Delta_t^{\tilde{d}} (y_t - \tilde{\mu}_T) 1\{t > 0\}. \tag{2.9}$$

Proposition 25 *a) Under Assumptions 1 and 2a or 2b and under $H_{1T}(\lambda_0)$, for the DGP (2.8), and for an unknown break fraction λ_0,*

$$\sup_{\lambda} LM_T(\lambda) \xrightarrow{d} \sup_{\lambda} \left\{ [\frac{\left(\lambda B(1) - B(\lambda) - \delta\left(\lambda(1-\lambda_0) - (\lambda-\lambda_0)_+\right)\omega\right)^2}{\lambda(1-\lambda)}] + [\frac{\left(\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}(1) - \tilde{W}_{1/2-d_0}(\lambda) - \eta\frac{\lambda^{1-2d_0}\left(1-\lambda_0^{1-2d_0}\right) - \left(\lambda^{1-2d_0}-\lambda_0^{1-2d_0}\right)_+}{\sqrt{1-2d_0}\Gamma(1-d_0)}\right)^2}{\lambda^{1-2d_0}\left(1-\lambda^{1-2d_0}\right)}] \right\},$$

b) Under Assumptions 1 and 2a or 2b and under $H_{1T}(\lambda_0)$, for a known break fraction λ_0,

$$LM_T \xrightarrow{d} \chi_2^2(c),$$

with non-centrality parameter

$$c = \delta^2\lambda_0(1-\lambda_0)\frac{\pi^2}{6}\omega^2 + \eta^2\frac{\lambda_0^{1-2d_0}\left(1-\lambda_0^{1-2d_0}\right)}{(1-2d_0)\Gamma^2(1-d_0)}.$$

Thus, the presence of a stable short run dynamics decreases the local power, i.e., it reduces ω^2, only through the component corresponding to the test for a break in the memory.

Changing autoregressive structure

Further, we can allow for changes in the autoregressive structure of the DGP, which so far have been ignored. In particular, we consider the process

$$\begin{array}{c} \alpha(L)\Delta_t^{d^0}(y_t-\mu_0) = \varepsilon_t,\ t=1,...,[\lambda_0 T] \\ [\alpha(L)+\beta(L)]\Delta_t^{d_0+\theta}(y_t-\mu_0-\nu) = \varepsilon_t,\ t=[\lambda_0 T]+1,...,T. \end{array} \tag{2.10}$$

Defining $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)$ and $\boldsymbol{\beta} = \left(\beta_1, ..., \beta_p\right)$, this leads to the following likelihood,

$$\begin{aligned} L\left(\theta, d, \boldsymbol{\beta}, \boldsymbol{\alpha}, \nu, \mu, \sigma^2, \lambda\right) &= -\frac{T}{2} \log\left(2\pi\sigma^2\right) \\ &\quad -\frac{1}{2\sigma^2} \left\{ \sum_{t=1}^{[\lambda T]} \alpha\left(L\right) \Delta_t^d \left(y_t - \mu\right) \right. \\ &\quad \left. + \sum_{t=[\lambda T]+1}^{T} \left[\alpha\left(L\right) + \beta\left(L\right)\right] \Delta_t^{d+\theta} \left(y_t - \mu - \nu\right) \right\}^2 . \end{aligned}$$

The derivative of the likelihood function in the direction of $\boldsymbol{\beta}$, evaluated at the restricted estimates $(0, \tilde{d}_T, 0, \tilde{\alpha}_T, 0, \tilde{\mu}_T, \tilde{\sigma}^2, \lambda)$, is given by,

$$\begin{aligned} \frac{\partial L\left(\theta, d, \boldsymbol{\beta}, \boldsymbol{\alpha}, \nu, \mu, \sigma^2, \lambda\right)}{\partial \boldsymbol{\beta}} &= \frac{1}{\tilde{\sigma}_T^2} \sum_{t=[\lambda T]+1}^{T} \tilde{\alpha}\left(L\right) \Delta_t^{\tilde{d}_T} \left(y_t - \tilde{\mu}\right) \begin{pmatrix} \Delta_{t-1}^{\tilde{d}_T} \left(y_{t-1} - \tilde{\mu}_T\right) \\ ... \\ \Delta_{t-p}^{\tilde{d}_T} \left(y_{t-p} - \tilde{\mu}_T\right) \end{pmatrix} \\ &= \frac{1}{\tilde{\sigma}_T^2} \sum_{t=[\lambda T]+1}^{T} \tilde{\varepsilon}_t \begin{pmatrix} \tilde{\alpha}_T^{-1}\left(L\right) \tilde{\varepsilon}_{t-1} \\ ... \\ \tilde{\alpha}_T^{-1}\left(L\right) \tilde{\varepsilon}_{t-p} \end{pmatrix} . \end{aligned}$$

As in Tanaka (1999), by a central limit theorem for martingale difference sequences, it follows that under the null,

$$\sqrt{T}\left(\tilde{d}_T, \tilde{\boldsymbol{\alpha}}_T'\right)' \xrightarrow{d} N\left(0, \Xi^{-1}\right),$$

where $\Xi = \begin{pmatrix} \pi^2/6 & \kappa \\ \kappa & \Phi \end{pmatrix}$, with Φ defined in Proposition 25. While the estimation of the memory and the short run dynamics are correlated, both are still asymptotically independent from the estimation of the level. Proposition 26 discusses the asymptotic behavior of a joint test for breaks in memory, level and short term dynamics under the local alternative

$$\left(\theta, \beta_1, ..., \beta_p, \nu\right) = \left(\delta/T^{1/2}, \boldsymbol{\gamma}'/T^{1/2}, \eta/T^{1/2-d_0}\right),$$

where $\boldsymbol{\gamma}' = \left(\gamma_1, ..., \gamma_p\right)$.

Proposition 26 *a) Under Assumptions 1 and 2a or 2b and under $H_{1T}\left(\lambda_0\right)$, for the*

DGP (2.10), for a unknown break fraction λ_0,

$$\sup_{\lambda} LM_T(\lambda) \xrightarrow{d} \sup_{\lambda} \left\{ \frac{\left(\lambda B_{p+1}(1) - B_{p+1}(\lambda) - \left(\lambda(1-\lambda_0) - (\lambda-\lambda_0)_+\right)\sqrt{\begin{pmatrix} \delta & \gamma' \end{pmatrix} \Xi \begin{pmatrix} \delta \\ \gamma \end{pmatrix}}\right)^2}{\lambda(1-\lambda)} \right.$$

$$\left. + \frac{\left(\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}(1) - \tilde{W}_{1/2-d_0}(\lambda) - \eta\frac{\lambda^{1-2d_0}\left(1-\lambda_0^{1-2d_0}\right) - \left(\lambda^{1-2d_0} - \lambda_0^{1-2d_0}\right)_+}{\sqrt{1-2d_0}\Gamma(1-d_0)}\right)^2}{\lambda^{1-2d_0}\left(1-\lambda^{1-2d_0}\right)} \right\}$$

where $B_{p+1}(.)$ is a $(p+1)-$dimensional Brownian motion.

b) Under Assumptions 1 and 2a or 2b and under H_{1T}, for the DGP (2.10), for a known break fraction λ_0,

$$LM_T \xrightarrow{d} \chi^2_{2+p}(c),$$

where

$$c = \lambda_0(1-\lambda_0)\begin{pmatrix} \delta & \gamma \end{pmatrix} \Xi \begin{pmatrix} \delta \\ \gamma \end{pmatrix} + \eta^2 \frac{\lambda_0^{1-2d_0}(1-\lambda_0^{1-2d_0})}{(1-2d_0)\Gamma^2(1-d_0)}.$$

Consistency of the LM test

Proposition 27 shows the consistency of the LM tests for either breaks in the level, memory or in both parameters simultaneously. In particular, we consider the following alternative hypotheses

$$\begin{array}{r} H_1^d(\lambda_0) : \theta \neq 0 \text{ and } \nu = 0 \\ H_1^\mu(\lambda_0) : \theta = 0 \text{ and } \nu \neq 0 \\ H_1^{d,\mu}(\lambda_0) : \theta \neq 0 \text{ and } \nu \neq 0 \end{array}.$$

Let

$$\bar{d} = \arg\min_{d} \left\{ \lambda_0 \frac{\Gamma(1-2(d_0-d))}{\Gamma^2(d-d_0+1)} + (1-\lambda_0)\frac{\Gamma(1-2(d_1-d))}{\Gamma^2(d-d_1+1)} \right\} \tag{2.11}$$

be the limit of the memory estimator under the alternative, with $\bar{d} > \lambda_0 d_0 + (1-\lambda_0) d_1$.

Two results stand out. First, the LM test statistics do not diverge asymptotically if the non- tested parameter is breaking. And secondly, the rate of divergence depends on which parameter is breaking, being T for a break only in d and T^{1-2d_0} for a break only in the level.

Proposition 27 *Under a fixed alternative of one break and for known and unknown break fraction*
a) Under Assumptions 2a and 2b, the LM tests for a break in both parameters, $LM_T^{d,\mu}(\lambda_0)$ *and* $\sup_\lambda LM_T^{d,\mu}(\lambda)$ *diverge at rate* T *under* $H_1^{d,\mu}(\lambda_0)$ *(breaks in memory and level) and under* $H_1^d(\lambda_0)$*, and they diverge at rate* T^{1-2d_0} *under* $H_1^\mu(\lambda_0)$*.*
b) The LM tests for a break in the memory, $LM_T^d(\lambda_0)$ *and* $\sup_\lambda LM_T^d(\lambda)$*, diverge at rate* T *under* $H_1^{d,\mu}$ *and* $H_1^d(\lambda_0)$*, and converge under* $H_1^\mu(\lambda_0)$ *to the term* $[I^d]$ *in Theorem 22 for* $\delta = 0$ *and to a* χ_1^2*, respectively. Under Assumption 2a and* $H_1^\mu(\lambda_0)$*, the LM tests converge to* $sup_\lambda[I^d(\lambda)]$ *with* $d = 0$*.*
c) The LM tests for a break in the level, $LM_T^\mu(\lambda_0)$ *and* $\sup_\lambda LM_T^\mu(\lambda)$ *diverge at rate* T^{1-2d_0} *under* $H_1^\mu(\lambda_0)$*, at rate* $T^{1-2\bar{d}}$ *under* $H_1^{d,\mu}(\lambda_0)$*, and converge under* H_1^d *to the term* $sup_\lambda[II^\mu(\lambda)]$ *in Theorem 22 for* $\eta = 0$ *and to a* χ_1^2 *respectively. Under Assumption 2a, the LM converges to the term* $sup_\lambda[II^\mu(\lambda)]$ *with* $\bar{d} = 0$*.*

2.2.2 Wald tests

Alternatively to the discussed LM test, we can derive a Wald type test along the lines of Dolado et al. (2002) and Lobato and Velasco (2007) who derive a Fractional Dickey Fuller test for the null hypothesis of a unit root process against the alternative of a process with a memory smaller than unity. Dolado et al. (2009) extend this test by allowing for testing any memory $d = d_0$ against the alternative $d \neq d_0$. They further consider the estimation of a deterministic component and show that preestimation of such a deterministic component does not affect the asymptotic distribution of the test. Further, they prove that, while remaining asymptotically equivalent under local alternatives, the EFDF test has considerably higher power than the LM under the alternative. In view of these results, we use this approach to construct a EFDF test for testing for breaks in the memory and in the level, in particular, for the hypothesis (H0) in model (2.2).

We construct an efficient test for breaks in memory and level in a similar way as the unit root test in in Lobato and Velasco (2007). We start with the expression under the alternative (maintained hypothesis),

$$\begin{cases} \Delta_t^{d_0}(y_t - \mu_0) = \varepsilon_t, \ t = 1, ..., [\lambda_0 T] \\ \Delta_t^{d_1}(y_t - \mu_1) = \varepsilon_t, \ t = [\lambda_0 T] + 1, ..., T. \end{cases}$$

For $t = [\lambda_0 T] + 1, ..., T$, this can be rewritten as,

$$\begin{aligned} \Delta_t^{d_0}(y_t - \mu_0) &= \Delta_t^{d_0}(y_t - \mu_0) - \Delta_t^{d_1}(y_t - \mu_1) + \varepsilon_t \\ &= (\Delta_t^{d_0} - \Delta_t^{d_1})(y_t - \mu_1) + \Delta_t^{d_0}(\mu_1 - \mu_0) + \varepsilon_t. \end{aligned}$$

Finally, using the same rationale as in Lobato and Velasco (2007) and recalling that $d_1 = d_0 + \theta$ and $\mu_1 = \mu_0 + \nu$, we run the following regression

$$\Delta_t^{d_0}(y_t - \mu_0) = \vartheta_1 \left[\frac{1 - \Delta_t^{\theta D_t(\lambda)}}{\theta} \right] \Delta_t^{d_0}(y_t - \mu_0 - \nu) + \vartheta_2 \Delta_t^{d_0} 1 D_t(\lambda) + \varepsilon_t, \tag{2.12}$$

where in practice, we need to estimate θ and ν. As pointed out by these authors, for $\theta \to 0$, the filter $\left[\frac{1-\Delta_t^{\theta}}{\theta}\right]$ becomes $\log \Delta$ which corresponds to the well-known lag filter used in the regression-based LM test.

Define $\Theta = (\vartheta_1, \vartheta_2)'$, $Y_t = \Delta_t^{d_0}(y_t - \mu_0)$, and

$$X_t(\lambda) = \left(\left[\frac{1 - \Delta_t^{\theta}}{\theta} \right] \Delta_t^{d_0}(y_t - \mu_0 - \nu) D_t(\lambda), \Delta_t^{d_0} 1 D_t(\lambda) \right)'.$$

Using this notation, (2.12) can be written as,

$$Y_t = \Theta X_t(\lambda) + \varepsilon_t.$$

Testing for breaks in both parameters corresponds to the joint null hypothesis of $\vartheta_1 = \vartheta_2 = 0$ in (2.12), while testing for a break only in the memory corresponds to the null hypothesis of $\vartheta_1 = 0$ and for a break only in the level, it corresponds to $\vartheta_2 = 0$.

First, we define the Wald test statistic as,

$$W_T(\lambda) = \hat{\Theta}(\lambda) \hat{V}^{-1}(\lambda) \hat{\Theta}(\lambda), \tag{2.13}$$

where, under homoskedasticity and a constant error variance,

$$\hat{V}(\lambda) = \hat{\sigma}^2 (X'X)^{-1}.$$

Next,

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^2$$

and

$$\hat{\varepsilon}_t = \begin{cases} \Delta_t^{d_0}(y_t - \mu_0), t = 1, ..., [\hat{\lambda}T] \\ \Delta_t^{d_1}(y_t - \mu_1), \ t = 1 + [\hat{\lambda}T], ..., T. \end{cases}$$

In the sequel, we analyze the behavior of the Wald test again under the local alternative $\mathrm{H}_{1T}(\lambda_0)$ in (2.7) for both a known and unknown break fraction. For the more realistic case of an unknown break fraction, we construct the Wald test statistic for a grid of values of the break fraction in order to choose the maximum,

$$\hat{\lambda} = \arg\max_{\lambda} W_T(\lambda).$$

Theorem 28 provides the asymptotic behavior of the Wald test for known and unknown break fraction

Theorem 28 *Under Assumptions 1 and 2a or 2b and under the local hypothesis* $H_{1T}(\lambda_0)$ *and*
a) an unknown break fraction λ_0*, the asymptotic behavior of the Wald test* $W_T(\lambda_0)$ *corresponds to the one of the LM test in Proposition 24 a).*
b) a known fixed break fraction λ_0*, the asymptotic behavior of the Wald test* $\sup_\lambda W_T(\lambda)$ *corresponds to the one of the LM test in Proposition 24 b).*

Theorem 28 nests again the cases of testing for a break only in the level and only in the memory. If we are sure that only one parameter is breaking, a testing procedure not allowing for a break in the non-tested parameter should lead to better finite sample properties (e.g. set $\mu_0 = \mu_1$ or $\nu = 0$ in (2.12) when testing for a break in the memory, $H_0 : \vartheta_1 = 0$).

So far, it has been assumed that the values of d_0 and μ_0 are known. From the discussion in Wooldridge (2002) and Lobato and Velasco (2007), it follows that the estimation of $\theta = d_1 - d_0$ and $\nu = \mu_1 - \mu_0$ does not affect the asymptotic distribution derived in Theorem 28. However, this is not true for the estimation of d_0 and μ_0 since these parameters affect the left-hand-side variable in the above-mentioned regression (2.12). Now, we relax this assumption by assuming that they are unknown and therefore need to be estimated before running the testing regression. In particular, given the break fraction λ, we estimate (d_0, μ_0) by the conditional sum of squares estimator, using the observations of the first regime, namely,

$$\left(\hat{d}_0(\lambda), \hat{\mu}_0(\lambda)\right) = \arg\min_{d_0,\mu_0} \sum_{t=1}^{[\lambda T]} \left(\Delta_t^{d_0}(y_t - \mu_0)\right)^2.$$

From Chapter 1, the resulting estimators behave as follows,

$$\begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{1/2-d_1} \end{pmatrix} \begin{pmatrix} \hat{d}_0(\lambda) - d_0 \\ \hat{\mu}_0(\lambda) - \mu_0 \end{pmatrix} \Longrightarrow \begin{pmatrix} \lambda^{-1}\frac{\sqrt{6}}{\pi} B(\lambda) \\ \sigma \frac{\Gamma(1-d_0)\sqrt{1-2d_0}}{\lambda^{1-2d_0}} \tilde{W}(\lambda) \end{pmatrix}. \tag{2.14}$$

Performing the Wald test using estimates $(\hat{d}_0, \hat{\mu}_0)$ in place of $(d_{0,}\mu_0)$, we get the following asymptotic distribution of the Wald test.

Theorem 29 *Under Assumptions 1 and 2a or 2b and under the local hypothesis H_{1T} and for unknown parameters d_0 and μ_0 and*
a) an unknown break fraction λ_0, the asymptotic behavior of the Wald test $W_T(\lambda_0)$ corresponds to the one of the LM test in Theorem 22.
b) a known break fraction λ_0, the asymptotic behavior of the Wald test $\sup_\lambda W_T(\lambda)$ corresponds to the one of the sup LM test in Corollary 23.

Stable short run dynamics

Next, we extend the previous results by considering the case of stable short run dynamics in equation (2.8). For estimating the parameters in the context of the unit root testing, Lobato and Velasco (2007) propose a two-step procedure. Dolado et al. (2009) advocate instead a one-step procedure that leads to the same asymptotic distribution. For simplicity, consider the case of a AR(1), with a stable coefficient, α. For the case of a non-breaking level, the one-step procedure leads to the following regression,

$$\Delta_t^{d_0}(y_t - \mu_0) = \vartheta_1 \left[\frac{1 - \Delta_t^{\theta D_t(\lambda)}}{\theta}\right] \Delta_t^{d_0}(y_t - \mu_0) - \alpha \Delta_{t-1}^{d_0+\theta D_t(\lambda)}(y_{t-1} - \mu_0) + \varepsilon_t.$$

As in the previous sections, testing for a break in the memory corresponds to testing $H_0 : \vartheta_1 = 0$ and its asymptotics corresponds to the first term in brackets in Proposition 2.

Next, when the level is also allowed to break, we start again with the DGP under the alternative,

$$\begin{aligned} (1-\alpha L)\Delta_t^{d_0}(y_t - \mu_0) &= \varepsilon_t, \ t = 1, ..., [\lambda_0 T] \\ (1-\alpha L)\Delta_t^{d_1}(y_t - \mu_1) &= \varepsilon_t, \ t = [\lambda_0 T] + 1, ..., T. \end{aligned}$$

Next, for $t = [\lambda_0 T] + 1, ..., T$,

$$\begin{aligned}
\Delta_t^{d_0}(y_t - \mu_0) &= \Delta_t^{d_0}(y_t - \mu_0) - (1 - \alpha L)\Delta_t^{d_1}(y_t - \mu_1) + \varepsilon_t \\
&= \left[\Delta_t^{d_0} - \Delta_t^{d_1}\right](y_t - \mu_1) + \Delta_t^{d_0}(\mu_0 - \mu_1) \\
&\quad + \alpha\Delta_{t-1}^{d_1}(y_{t-1} - \mu_1) + \varepsilon_t.
\end{aligned}$$

The first term behaves the same as in the case of a non-breaking level. The second term is prespecified since it is deterministic and the third term is lagged. Hence, we can estimate by OLS the following regression model

$$\begin{aligned}
\Delta_t^{d_0}(y_t - \mu_0) &= \vartheta_1\left[\frac{1 - \Delta_t^{\theta D_t(\lambda)}}{\theta}\right]\Delta_t^{d_0}(y_t - \mu_0 - \nu) + \vartheta_2 D_t(\lambda)\Delta_t^{d_0}1 \\
&\quad + \alpha\left[\Delta_{t-1}^{d_0+\theta D_t(\lambda)}(y_{t-1} - \mu_0) - D_t(\lambda)\Delta_{t-1}^{d_0+\theta}\nu\right] + \varepsilon_t.
\end{aligned}$$

Testing for a break in both parameters corresponds to testing

$$H_0 : \vartheta_1 = \vartheta_2 = 0.$$

Proposition 30 provides the asymptotic behavior of the Wald test for this case.

Proposition 30 *For the DGP (2.10), for a unknown break fraction, and Assumptions 1 and 2a or 2b*
a) Under H_{1T}, for a unknown break fraction, $\sup_\lambda W_T(\lambda)$ behaves asymptotically as the LM test in Proposition 25 a).
b) Under H_{1T}, for a known break fraction, $W_T(\lambda)$ behaves asymptotically as the LM test in Proposition 25 b).

Changing short run dynamics

Similarly, we can modify the Wald regression in order to allow for breaks in the short run dynamics. In particular, modify the previous regression model by

$$\begin{aligned}
\Delta_t^{d_0}(y_t - \mu_0) &= \vartheta_1\left[\frac{1 - \Delta_t^{\theta D_t(\lambda)}}{\theta}\right]\Delta_t^{d_0}(y_t - \mu_0 - \nu) + \vartheta_2 D_t(\lambda)\Delta_t^{d_0}1 \\
&\quad + (\alpha + \beta D_t(\lambda))\left[\Delta_{t-1}^{d_0+\theta D_t(\lambda)}(y_{t-1} - \mu_0) - D_t(\lambda)\Delta_{t-1}^{d_0+\theta}\nu\right] + \varepsilon_t.
\end{aligned}$$

Now, a test for

$$H_0 : \vartheta_1 = \vartheta_2 = 0.$$

is robust to potential breaks in the short run dynamics.

Consistency

Next, Proposition 31 discusses the consistency of the Wald test, under fixed alternatives, for breaks in both parameters. If we know that one parameter is not breaking, and we wish to test whether the other parameter is breaking, the power of the Wald test can be improved by testing exclusively for a break in the latter.

Proposition 31 *a) The Wald tests for a break in both parameters,* $W_T^{d,\mu}(\lambda_0)$ *and* $\sup_\lambda W_T^{d,\mu}(\lambda)$, *behave as the LM test in Proposition 27 a).*
b) Under Assumption 2b, the Wald tests for a break in the memory, $W_T^{d}(\lambda_0)$ *and* $\sup_\lambda W_T^{d}(\lambda)$, *behave as the LM test in Proposition 27 b). Under Assumption 2a and under* H_1^μ, *the Wald tests converge to* $\sup_\lambda[I^d(\lambda)]$, *with* $d{=}d_0 \leq 0$.
c) Under Assumption 2a and b, the Wald tests for a break in the level, $W_T^{\mu}(\lambda_0)$ *and* $\sup_\lambda W_T^{\mu}(\lambda)$, *behave as the LM test in Proposition 27 c).*

First, note the Wald test is consistent for breaks of the memory in both directions. This contrasts with the CUSUM estimator of Kim et al. (2002), which has to be adjusted to be consistent for breaks in both directions.

2.3 Comparing the behavior of the tests under the alternative of one break

In Chapter 1, I propose an alternative sup F-test ($supF$) for the problem of testing for breaks in the level and/or the memory. In the preceding analysis, we have shown that, under the considered local alternatives, the asymptotic distributions of the LM and the Wald test are asymptotically equivalent. Further, from Chapter 1, the F-test has the same asymptotic distribution as well. However, the same is not true under the fixed alternative hypothesis. It is well known that Wald type tests have higher power than the LR test and that this one beats the LM test in this respect. For fractional unit root tests, Lobato and Velasco (2007) have confirmed this result through simulations, and Dolado et al (2007) and Lobato and Velasco (2008) have proven it analytically. The latter consider the probabilistic limit of the properly normalized test statistic in a fixed-alternative framework.

2.3.1 Power performance of LM, Wald and F tests

First, we consider fixed alternatives of breaks in memory and/or level. We assume the DGP in (2.1). We compare the LM and Wald test to a LR test, in terms of the supF-test proposed in Chapter 1. In the notation of Chapter 1,

$$F(\lambda) = \frac{SSR_0 - SSR_1(\lambda)}{\frac{1}{T} SSR_1(\lambda)}, \tag{2.15}$$

where SSR_0 denotes the sum of squared errors under the null and $SSR_1(\lambda)$ denotes the sum of squared errors under the alternative of one break.

In the sequel, we consider the following three relevant cases: (i) a break only in the memory (bd), (ii) a break only in the level (bl), and (iii) a break in both parameters (bdl). As regards case (bd), following Lobato and Velasco (2008), Proposition 32 provides the power for the three tests - LM, Wald and F-test.

The following proposition characterizes the limiting behavior of the three test statistics under (bd).

Proposition 32 *Under the fixed alternative (2.1) with one break in the memory ($d_0 \neq d_1, \mu_0 = \mu_1$),*
a) The LM test (2.4) satisfies

$$p \lim_{T \to \infty} \frac{\sup_\lambda LM}{T} = (1 - \lambda_0) \frac{6}{\pi^2} \frac{\Gamma^2\left(1 + \bar{d} - d_1\right)}{\Gamma^2\left(d_1 - \bar{d}\right)} \left(\sum_{j=1}^{\infty} \frac{\Gamma\left(j + d_1 - \bar{d}\right)}{j \Gamma\left(j + 1 + \bar{d} - d_1\right)} \right)^2.$$

b) The Wald test (2.13) satisfies

$$p \lim_{T \to \infty} \frac{\sup_\lambda W}{T} = (1 - \lambda_0) \sum_{j=1}^{\infty} \pi_j^2 (d_0 - d_1) = (1 - \lambda_0) \left[\frac{\Gamma(1 + 2(d_0 - d_1))}{\Gamma^2(1 + (d_0 - d_1))} - 1 \right].$$

c) The F-test (2.15) satisfies

$$p \lim_{T \to \infty} \frac{\sup_\lambda F}{T} = (1 - \lambda_0) \log \left(\frac{\Gamma\left(1 + 2\left(\bar{d} - d_1\right)\right)}{\Gamma^2\left(1 + \left(\bar{d} - d_1\right)\right)} \right),$$

where $\bar{d}$ is defined in (2.11).

Note that for the LM test, the power differs between the cases when we estimate d_0 using the whole sample or when it is assumed to be known. Not surprisingly, in the latter case the power is larger. Next, Proposition 33 provides the power for the case (bl).

Proposition 33 *Under the fixed alternative (2.1) with one break in the level ($d_0 = d_1, \mu_0 \neq \mu_1$),* *a) the LM test (2.4) satisfies*

$$\underset{T\to\infty}{p\lim} T^{2d_0-1} \sup_{\lambda} LM = \left(\frac{\mu_1-\mu_0}{\sigma}\right)^2 \frac{\lambda_0^{1-2d_0}(1-\lambda_0^{1-2d_0})}{(1-2d_0)\,\Gamma^2(1-d_0)}. \tag{2.16}$$

b) the Wald test (2.13) satisfies

$$\underset{T\to\infty}{p\lim} T^{2d_0-1} \sup_{\lambda} W = \left(\frac{\mu_1-\mu_0}{\sigma}\right)^2 \frac{(1-\lambda_0^{1-2d_0})}{(1-2d_0)\,\Gamma^2(1-d_0)}. \tag{2.17}$$

c) the F-test (2.15) satisfies

$$\underset{T\to\infty}{p\lim} T^{2d_0-1} \sup_{\lambda} F = \left(\frac{\mu_1-\mu_0}{\sigma}\right)^2 \frac{\lambda^{1-2d_0}(1-\lambda_0^{1-2d_0})}{(1-2d_0)\,\Gamma^2(1-d_0)}. \tag{2.18}$$

Note that the result in (2.17) does not change if μ_0 is estimated since, from the proof of Theorem 28, we have that

$$\hat{\vartheta}_2 = \frac{T^{2d_0-1}\sum_{t=[\lambda T]+1}^{T} \Delta_t^{d_0}1\Delta_t^{d_0}(y_t-\mu_0)}{T^{2d_0-1}\sum_{t=[\lambda T]+1}^{T}(\Delta_t^{d_0}1)^2} - (\hat{\mu}_0-\mu_0),$$

where the first term in the right hand side behaves as in (2.17) while the second term converges in probability to *zero.*

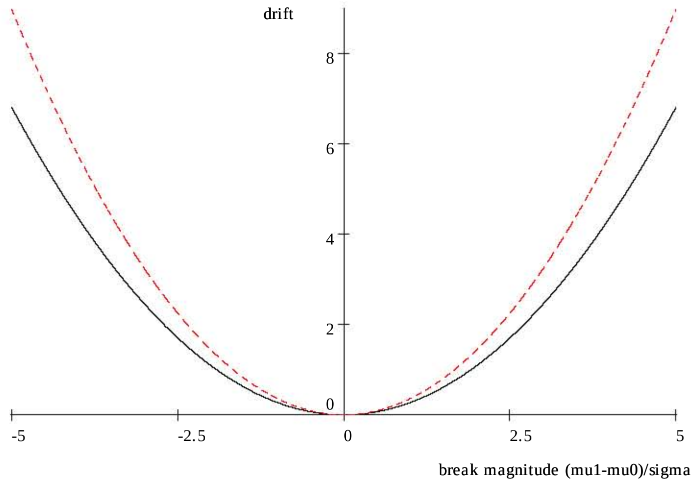
Thus, a comparison of the supLM test and the supF-test for case (i) shows that, interestingly, both exhibit the same asymptotic power. The reason is that, unlike in the usual cases, for (bl) the variance is equally efficiently estimated under the null as under the alternative $H_1^{\mu}(\lambda_0)$.

Finally, in case (bdl), the break in the memory dominates for $0 \leq d_0, d_1 < 1/2$ and $T^{-1}\sup LM, T^{-1}\sup W$ and $T^{-1}\sup F$ behave as in Proposition 32. Comparing (2.17) and (2.16) under fixed alternatives, we can see that, similarly to what happens in the direction of the memory, similarly as discussed in Lobato and Velasco (2008), the supWald test is more powerful than the supLM test in the direction of the level.

2.3.2 Power comparison under fixed alternatives

In this section, we plot the non-centrality parameters of the tests for case (bl), when only the level is breaking. We do this for the power of the LM/F-tests, expression (2.16), (black solid line) and the power for the Wald test, expression (2.17), (red dashed line). Since, the non-centrality parameters depend on the memory (d_0),

Figure 2-2: **Drift of the tests as a function of the break magnitude**



LM test and F-test (black solid line) and Wald test (red dashed line). λ_0= 0.5 and $d_0 = 0.3$.

on the break fraction (λ) and on the size of the break in the level ($\frac{\mu_1 - \mu_0}{\sigma}$), we respectively vary one parameter at a time, while keeping the other two parameters fixed. Figure 2-2 shows the power function regarding different break magnitudes for $\lambda = 0.5, d = 0.3$ and $\mu_0 = 0$.

Not surprisingly, the drift is increasing and symmetric in the break magnitude $\frac{\mu_1 - \mu_0}{\sigma}$. Likewise, Figure 2-3 displays the efficiency of the three tests regarding different break fractions for $d = 0.3$ and $\mu - \mu_0 = 1$.

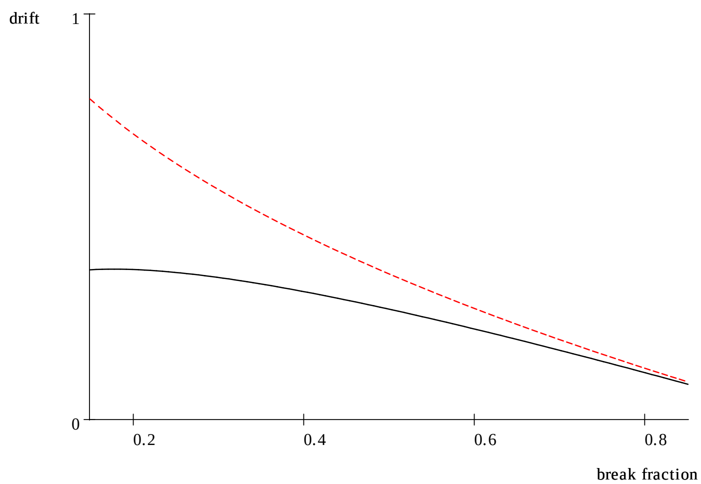
Notice that the power is not symmetric around $\lambda = 0.5$. First, as seen in Figure 2-1, λ^{1-2d_0} $(1 - \lambda^{1-2d_0})$ is not symmetric. Second, in the Wald test, we estimate consistently d_0, for any $[\lambda T]$. Hence, the power is increasing in $(1 - \lambda)$. Finally, Figure 2-5 displays the efficiency of the three tests for different values of the memory parameter for $\lambda = 0.5$ and $\mu_2^0 - \mu_1^0 = 1$. Notice that the relative power of the Wald test is decreasing in the memory parameter d_0.

For the break in the memory, the relative powers resemble Figure 1 in Lobato and Velasco (2008) multiplied by $(1 - \lambda)$ (see Figure 2-5).

The drift is much larger if the memory decreases than it is if the memory increases by the same amount.

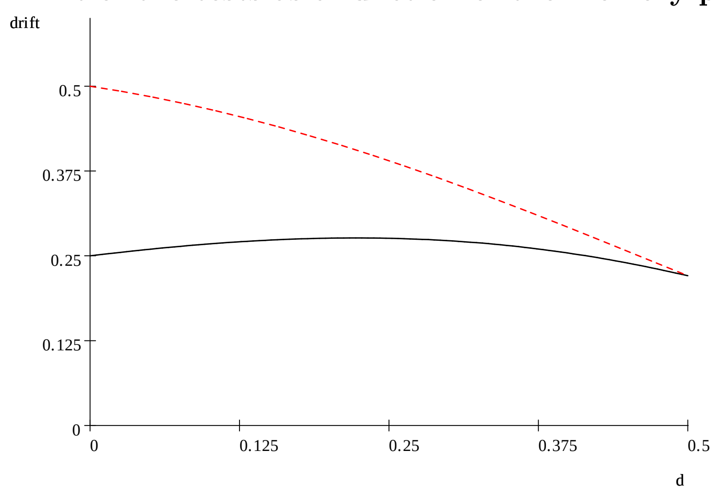
Overall, the results in this section show that the Wald test exhibits better power properties under fixed alternatives than the LM and F-tests.

Figure 2-3: **Drift of the tests as a function of the break fraction**



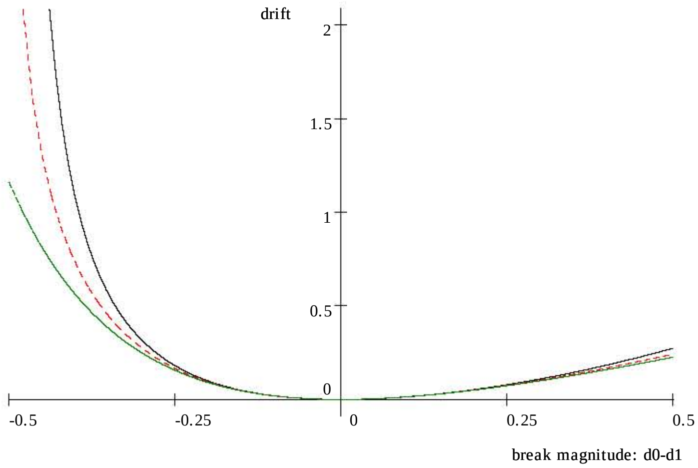
LM test and F-test (black solid line) and Wald test (red dashed line). $d = 0.3$ and $\left|\frac{\mu_1 - \mu_0}{\sigma}\right| = 1.$

Figure 2-4: **Drift of the tests as a function of the memory parameter**



LM test and F-test (black solid line) and Wald test (red dashed line). $\lambda_0 = 0.5$, $\left|\frac{\mu_1 - \mu_0}{\sigma}\right| = 1.$

Figure 2-5: **Drift of the tests as a function of the break magnitude**



LM test (green line), F-test (red dashed line) and Wald test (black solid line). $\lambda_0 = 0.5$.

2.4 Final Remarks

The proposed methodology can be extended along the lines in Chapter 1 to the presence of multiple breaks. Consider for $i = 0, ..., m-1$

$$\Delta_t^{d_i}(y_t - \mu_i) = \varepsilon_t,\ t = [\lambda_{i-1}T] + 1, ..., [\lambda_i T].$$

Consider the case of testing for 0 vs 2 breaks. The LM test uses the following likelihood function

$$\begin{aligned} & L\left(\theta_1, \theta_2, d, \nu_1, \nu_2, \mu, \sigma^2, \lambda_1, \lambda_2\right) \\ = & -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left\{\Delta_t^{d+\theta_1 D_t^1(\lambda)}(y_t - \mu) - D_t^1(\lambda)\,\Delta_t^{d+\theta_1 D_t(\lambda)}\nu_1\right. \\ & \left. +\Delta_t^{d+\theta_2 D_t^2(\lambda)}(y_t - \mu) - D_t^2(\lambda)\,\Delta_t^{d+\theta_2 D_t(\lambda)}\nu_2\right\}^2, \end{aligned}$$

where $D_t^1(\lambda) = 1([\lambda_1 T] < t \leq [\lambda_2 T])$ and $D_t^2(\lambda) = 1(t > [\lambda_2 T])$. The test is constructed as in (2.4). The asymptotic distribution is a sum of terms as in Theorem 1. Again, we can construct a Wald test running the regression

$$\begin{aligned} \Delta_t^{d_0}(y_t - \mu_0) = & \left[\vartheta_1\left[\frac{1-\Delta_t^{\theta_1}}{\theta_1}\right]\Delta_t^{d_0}(y_t - \mu_0 - \nu_1) + \vartheta_2\Delta_t^{d_0}1\right]D_t^1(\lambda) \\ & + \left[\vartheta_3\left[\frac{1-\Delta_t^{\theta_2}}{\theta_2}\right]\Delta_t^{d_0}(y_t - \mu_0 - \nu_2) + \vartheta_4\Delta_t^{d_0}1\right]D_t^2(\lambda) + \varepsilon_t. \end{aligned}$$

In this setup, similarly to Chapter 1, we can allow for non simultaneous breaks. Further, it is possible to construct tests that allow us to determine which is the changing parameter. In particular, for the LM test, we allow for a break in the non-tested parameter in order to obtain the residuals under the new null rather than (2.5). While for the Wald test, by allowing for a break in the non-tested parameter in (2.12), we are able to control for spurious effects stemming from that parameter (see the discussion in Chapter 1).

In Chapter 1, I show that the alternative sup F test procedure suffers from size distortions and advocate the use of a Bootstrap procedure. Therefore, in practice, we suggest using a similar Bootstrap procedure for the LM test and Wald test. These procedures can be designed to test breaks in one or in both parameters. Finally, again similarly to Chapter 1, we can construct bootstrap tests for breaks in memory (level) robust to a change in the level (memory).

In summary, first, we have shown the importance of joint modeling of breaks in the memory and the level. By considering both, we are able to avoid a potential confounding problem. Second, the considered tests have several advantages. LM tests are computationally attractive because they require only estimation under the null, while Wald tests can exploit further information on the alternative, potentially leading to higher power. Thus, comparing the three different tests, there are potentially gains from using the Wald test.

2.5 Appendix

Proof of Theorem 22

Assuming a unknown break fraction λ, under the local alternative H_{1T}, the LM test is based on the derivatives of L in direction of θ and ν, evaluated at the restricted estimates $(0, \tilde{d}_T, 0, \tilde{\mu}_T, 0, \tilde{\sigma}_T^2)$, given λ.

First, in the direction of θ,

$$\begin{aligned}
\widetilde{LM}_{\theta,T}(\lambda) &= \frac{\partial}{\partial\theta} L\left(\theta, d, \nu, \mu, \sigma^2, \lambda\right)\Big|_{\theta=0, d=\tilde{d}_T, \nu=0, \mu=\tilde{\mu}_T, \sigma^2=\tilde{\sigma}_T^2} \\
&= -\frac{1}{\tilde{\sigma}_T^2} \sum_{t=1}^{T} \left\{ D_t(\lambda) \log \Delta \Delta_t^{\tilde{d}} (y_t - \tilde{\mu}) \right\} \left\{ \Delta_t^{\tilde{d}} (y_t - \tilde{\mu}) \right\}
\end{aligned}$$

where

$$\left(\tilde{d}_T, \tilde{\mu}_T, \tilde{\sigma}_T^2\right) = \arg\max_{d,\mu,\sigma^2} L\left(0, d, 0, \mu, \sigma^2\right)$$

are the usual ML estimates without break. The consistency and rate of this estimator follows for Assumption 2b from Chapter 1. For Assumption 2a it behaves correspondingly.

Next, the usual ML residuals $\tilde{\varepsilon}_t$ are given by

$$\tilde{\varepsilon}_t = \Delta_t^{\tilde{d}} (y_t - \tilde{\mu}_T) 1\{t > 0\}$$

and $\tilde{\sigma}_T^2 = T^{-1} \sum_{t=1}^{T} \tilde{\varepsilon}_t^2$. Then, we have that,

$$\begin{aligned}
\widetilde{LM}_{\theta,T}(\lambda) &= -\frac{1}{\tilde{\sigma}_T^2} \sum_{t=[\lambda T]+1}^{T} \{\log \Delta \tilde{\varepsilon}_t\} \tilde{\varepsilon}_t = \frac{1}{\tilde{\sigma}_T^2} \sum_{t=[\lambda T]+1}^{T} \left\{ \sum_{k=1}^{t-1} \frac{1}{k} \tilde{\varepsilon}_{t-k} \right\} \tilde{\varepsilon}_t \\
&= \frac{1}{\tilde{\sigma}_T^2} \sum_{k=1}^{T-1} \frac{1}{k} \sum_{t=\max\{[\lambda T]+1, k+1\}}^{T} \tilde{\varepsilon}_t \tilde{\varepsilon}_{t+k} = \frac{1}{\tilde{\sigma}_T^2} \sum_{k=1}^{T-1} \frac{1}{k} \sum_{t=\max\{[\lambda T]+1-k, 1\}}^{T-k} \tilde{\varepsilon}_t \tilde{\varepsilon}_{t+k} \\
&= T \sum_{k=1}^{T-1} \frac{1}{k} \tilde{\rho}_k^*(\tilde{\varepsilon}_t),
\end{aligned}$$

where

$$\tilde{\rho}_k^*(\tilde{\varepsilon}_t) = \frac{\tilde{\sigma}_T^{-2}}{T} \sum_{t=\max\{\lambda T+1-k, 1\}}^{T-k} \tilde{\varepsilon}_t \tilde{\varepsilon}_{t+k}.$$

Under H_{1T},

$$T^{1/2} \left(\tilde{\rho}_1^*(\varepsilon_t(\theta_0)), \ldots, \tilde{\rho}_k^*(\varepsilon_t(\theta_0))\right) \Longrightarrow \left\{ B_j(1) - B_j(\lambda) - \delta j^{-1} \right\}_{j=1,\ldots,k}.$$

Now we approximate

$$\tilde{\rho}_k^*(\tilde{\varepsilon}_t) = \tilde{\rho}_k^*(\varepsilon_t(\theta_0)) + (\tilde{d}_T - d_0) \tilde{\rho}_{d,k}^{*\prime}(\varepsilon_t(\theta_0)) + (\tilde{\mu}_T - \mu_0) \tilde{\rho}_{\mu,k}^{*\prime}(\varepsilon_t(\theta_0)) + o_p\left(T^{-1/2}\right) \quad (2.19)$$

where, $k = 1, 2, \ldots$, fixed with T, and

$$\begin{aligned} \varepsilon_t(\theta_0, \nu_0) &= \Delta^{-\theta_0 D_t(\lambda)} \varepsilon_t - D_t(\lambda) \nu_0 \Delta_t^{d_0} 1 \\ &\approx \left(1 - \delta / T^{1/2} D_t(\lambda) \log \Delta\right) \varepsilon_t - \eta / T^{1/2-d_0} D_t(\lambda) \Delta_t^{d_0} 1. \end{aligned} \tag{2.20}$$

Next,

$$\begin{aligned} \tilde{\rho}_k^{*\prime}(\varepsilon_t(\theta_0)) &= \frac{\tilde{\sigma}_T^{-2}}{T} \sum_{t=\max\{[\lambda T]+1-k,1\}}^{T-k} \left(\log \Delta \varepsilon_t(\theta_0) \varepsilon_{t+k}(\theta_0) + \varepsilon_t(\theta_0) \log \Delta \varepsilon_{t+k}(\theta_0)\right) \\ &\xrightarrow{p} (1-\lambda) \frac{1}{k}, \end{aligned} \tag{2.21}$$

and

$$T^{1/2}\left(\tilde{d}_T - d_0\right) = -T^{1/2}\left(\frac{\pi^2}{6}\right)^{-1} \sum_{k=1}^{T-1} \frac{1}{k} \tilde{\rho}_k(\varepsilon_t(\theta_0)) + o_p(1),$$

where

$$\tilde{\rho}_k((\theta_0)) = \frac{\tilde{\sigma}_T^{-2}}{T} \sum_{t=1}^{T-k} \varepsilon_t(\theta_0) \varepsilon_{t+k}(\theta_0)$$

does not depend on λ.

Therefore from (2.19)

$$\begin{aligned} T^{-1/2} LM_{\theta,T}(\lambda) &= T^{1/2} \sum_{k=1}^{T-1} \frac{1}{k} \tilde{\rho}_k^*(\hat{\varepsilon}_t) \\ &= T^{1/2} \sum_{k=1}^{T-1} \left\{ \frac{1}{k} \tilde{\rho}_k^*(\varepsilon_t(\theta_0)) + \frac{1}{k^2}(1-\lambda)\left(\tilde{d}_T - d_0\right) \right\} + o_p(1) \quad (2.22) \\ &= T^{1/2} \sum_{k=1}^{T-1} \frac{1}{k} \tilde{\rho}_k^*(\varepsilon_t(\theta_0)) + T^{1/2}(1-\lambda)\left(\tilde{d}_T - d_0\right) \frac{\pi^2}{6} + o_p(1) \\ &= T^{1/2} \sum_{k=1}^{T-1} \frac{1}{k} \left\{ \tilde{\rho}_k^*(\varepsilon_t(\theta_0)) - (1-\lambda) \tilde{\rho}_k(\varepsilon_t(\theta_0)) \right\} + o_p(1) \\ &= T^{1/2} \sum_{k=1}^{T-1} \frac{1}{k} \breve{\rho}_k^*(\varepsilon_t(\theta_0)) + o_p(1), \end{aligned}$$

where

$$\begin{aligned} \breve{\rho}_k^*(\varepsilon_t(\theta_0)) &= \frac{\tilde{\sigma}_T^{-2}}{T} \sum_{t=\max\{\lambda T+1-k,1\}}^{T-k} \varepsilon_t(\theta_0) \varepsilon_{t+k}(\theta_0) \\ &\quad - (1-\lambda) \frac{\tilde{\sigma}_T^{-2}}{T} \sum_{t=1}^{T-k} \varepsilon_t(\theta_0) \varepsilon_{t+k}(\theta_0) + o_p(1). \end{aligned} \tag{2.23}$$

Under H_{1T},

$$
\begin{aligned}
& T^{1/2}\left(\check{\rho}_1^*\left(\varepsilon_t\left(\theta_0\right)\right),\ldots,\check{\rho}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)\right) && (2.24)\\
\Longrightarrow\ & \left\{B_j\left(1\right)-B_j\left(\lambda\right)+\delta\left(1-\max\{\lambda,\lambda_0\}\right)j^{-1}-\left(1-\lambda\right)\left[B_j\left(1\right)+\delta\left(1-\lambda_0\right)j^{-1}\right]\right\}_{j=1,\ldots,k}\\
=\ & \left\{\lambda B_j\left(1\right)-B_j\left(\lambda\right)-\delta\left(\lambda\left(1-\lambda_0\right)-\left(\lambda-\lambda_0\right)_+\right)j^{-1}\right\}_{j=1,\ldots,k}
\end{aligned}
$$

following from a standard FCLT and because the first term of $T^{1/2}\check{\rho}_j^*\left(\varepsilon_t\left(\theta_0\right)\right)$ has a drift of $\delta\left(1-\max\{\lambda,\lambda_0\}\right)j^{-1}$, while the second one has one of $\delta\left(1-\lambda\right)\left(1-\lambda_0\right)j^{-1}$; the latter since $T^{1/2}(\tilde{d}_T-d_0)$ has a drift of $\left(1-\lambda_0\right)\delta$ under H_{1T}.

Then, we can show that under H_{1T},

$$
\begin{aligned}
T^{-1/2}\widetilde{LM}_{\theta,T}\left(\lambda\right) &= T^{1/2}\sum_{k=1}^{T-1}\frac{1}{k}\check{\rho}_k^*\left(\varepsilon_t\left(\theta_0\right),\lambda\right)+o_p\left(1\right)\\
&\Longrightarrow -\delta\left(\lambda\left(1-\lambda_0\right)-\left(\lambda-\lambda_0\right)_+\right)\frac{\pi^2}{6}+\frac{\pi}{\sqrt{6}}\left(\lambda B\left(1\right)-B\left(\lambda\right)\right). \quad (2.25)
\end{aligned}
$$

In particular, let

$$
\begin{aligned}
Y_T &= T^{1/2}\sum_{k=1}^{T-1}\frac{1}{k}\check{\rho}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)\ \text{and}\\
X &= \frac{\pi}{\sqrt{6}}\left(\lambda B\left(1\right)-B\left(\lambda\right)\right).
\end{aligned}
$$

In order to show that $Y_T\Longrightarrow X$, we first define

$$
\begin{aligned}
X_{uT} &= T^{1/2}\sum_{k=1}^{u}\frac{1}{k}\check{\rho}_{T,k}^*\left(\varepsilon_t\left(\theta_0\right)\right)\ \text{and}\\
X_u &= \sum_{k=1}^{u}k^{-1}\left[\left(\lambda B_k\left(1\right)-B_k\left(\lambda\right)\right)\right],
\end{aligned}
$$

with the index T reflecting the dependence of $\check{\rho}_{T,k}^*$ on T. From Theorem 4.2 of Billingsley (1968), we have to show

1) for each u, $X_{uT}\Longrightarrow X_u$ as $T\to\infty$.
2) $X_u\Longrightarrow X$ as $u\to\infty$.
3) $\lim_{u\to\infty,T\to\infty}\sup P\{\sup_\lambda\left|X_{uT}-Y_T\right|\geq\varepsilon\}=0$.

1. From (2.24) and from a continuous mapping theorem, for a finite u,

$$
T^{1/2}\sum_{k=1}^{u}\frac{1}{k}\check{\rho}_{T,k}^*\left(\varepsilon_t\left(\theta_0\right)\right)\Longrightarrow\sum_{k=1}^{u}\frac{1}{k}\left[\left(\lambda B_k\left(1\right)-B_k\left(\lambda\right)\right)\right]
$$

2. For $u \to \infty$,

$$\sum_{k=1}^{u} \frac{1}{k} \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right] \Longrightarrow \sum_{k=1}^{\infty} \frac{1}{k} \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right], \tag{2.26}$$

which is the same process as X since it is Gaussian and has the same mean and the same variance. For the latter, using the uncorrelatedness of the covariance functions,

$$\begin{aligned} E\left[\sum_{k=1}^{\infty} \frac{1}{k} \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right] \right]^2 &= \sum_{k=1}^{\infty} k^{-2} E \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right]^2 \\ &= \frac{\pi^2}{6} \left[\lambda^2 + \lambda - 2\lambda^2 \right] = \frac{\pi^2}{6} \lambda (1 - \lambda) \end{aligned}$$

For proving (2.26), we have to show that uniformly in λ,

$$\sum_{k=u+1}^{\infty} \frac{1}{k} \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right] \xrightarrow{p} 0,$$

which follows from

$$\lim_{u \to \infty} \sup_{\lambda} \left| \sum_{k=u+1}^{\infty} \frac{1}{k} \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right] \right| \xrightarrow{p} 0.$$

Thus it suffices to show (mean square convergence)

$$\begin{aligned} & \lim_{u \to \infty} E \left[\sup_{\lambda} \left| \sum_{k=u+1}^{\infty} \frac{1}{k} \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right] \right| \right]^2 \\ = \; & \lim_{u \to \infty} E \sup_{\lambda} \left[\sum_{k=u+1}^{\infty} \frac{1}{k} \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right] \right]^2 \\ \leq \; & \lim_{u \to \infty} \sum_{k=u+1}^{\infty} \frac{1}{k^2} E \sup_{\lambda} \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right]^2 \qquad (2.27) \\ & + \lim_{u \to \infty} \sum_{k=u+1}^{\infty} \sum_{l=u+1}^{\infty} E \sup_{\lambda} \frac{1}{kl} \left[\left(\lambda B_k(1) - B_k(\lambda) \right) \right] \left[\left(\lambda B_l(1) - B_l(\lambda) \right) \right] \to 0. \end{aligned}$$

First, we use that for S_0 being the supremum of a Brownian Bridge,

$$S_0^2 \overset{d}{=} \frac{1}{2} e,$$

where e is a standard exponential distribution. Thus, $ES_0^2 = 1/2$ and (2.27) converges to *zero*. Next, using the Cauchy Schwarz inequality, convergence of the second term follows from showing it for the first term.

3. To show

$$\lim_{u\to\infty,T\to\infty}\sup P\{\sup_{\lambda}|\sum_{k=u}^{T}\frac{1}{k}\check{\rho}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)|\geq\varepsilon\}=0,$$

we first write $\check{\rho}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)=\frac{\gamma_k^*(\varepsilon_t(\theta_0))}{\gamma_0^*(\varepsilon_t(\theta_0))}$ where we rewrite the probability as

$$P\{\sup_{\lambda}|\sum_{k=u}^{T}\frac{1}{k}\gamma_k^*\left(\varepsilon_t\left(\theta_0\right)\right)|\geq\frac{\varepsilon\sigma^2}{2}\},$$

which converges to zero when $P\{\sup_\lambda|\sum_{k=u}^{T}\frac{1}{k}\gamma_k^*\left(\varepsilon_t\left(\theta_0\right)\right)|\geq\varepsilon\}$ does. Using the Markov inequality,

$$P\{\sup_{\lambda}|\sum_{k=u}^{T}\frac{1}{k}\check{\gamma}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)|\geq\varepsilon\}\leq\frac{E\sup_\lambda|\sum_{k=u}^{T}\frac{1}{k}\check{\gamma}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)|}{\varepsilon}.$$

Thus it suffices to show (mean square convergence) that

$$\begin{aligned}E\left(\sup_{\lambda}|\sum_{k=u}^{T}\frac{1}{k}\check{\gamma}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)|\right)^2 &= E\sup_{\lambda}\left(|\sum_{k=u}^{T}\frac{1}{k}\check{\gamma}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)|\right)^2\\ \leq\quad E\sup_{\lambda}\sum_{k=u}^{T}\frac{1}{k^2}\check{\gamma}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)^2&+E\sup_{\lambda}\sum_{k=u}^{T}\sum_{l=u}^{T}\frac{1}{kl}\check{\gamma}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)\check{\gamma}_l^*\left(\varepsilon_t\left(\theta_0\right)\right)\to 0\end{aligned}\qquad(2.28)$$

For the first term substituting the definition (2.23),

$$\begin{aligned}E\sup_{\lambda}\sum_{k=u}^{T}\frac{1}{k^2}\check{\gamma}_k^*\left(\varepsilon_t\left(\theta_0\right)\right)^2 \quad\leq\quad & E\sup_{\lambda}\sum_{k=u}^{T}\frac{1}{k^2}\left(\frac{1}{T^{1/2}}\sum_{t=\max\{[\lambda T]+1-k,1\}}^{T-k}\varepsilon_t\left(\theta_0\right)\varepsilon_{t+k}\left(\theta_0\right)\right)^2\\ &+E\sup_{\lambda}\sum_{k=u}^{T}\frac{1}{k^2}\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T-k}\varepsilon_t\left(\theta_0\right)\varepsilon_{t+k}\left(\theta_0\right)(1-\lambda)\right)^2.\end{aligned}$$

For the first term, first for any λ,

$$\begin{aligned}&E\sum_{k=u}^{T}\frac{1}{k^2}\left(\frac{1}{T^{1/2}}\sum_{t=\max\{[\lambda T]+1-k,1\}}^{T-k}\varepsilon_t\left(\theta_0\right)\varepsilon_{t+k}\left(\theta_0\right)\right)^2\\ =\quad&\sum_{k=u}^{T}\frac{1}{k^2}\frac{1}{T}\left(E\sum_{t=\max\{[\lambda T]+1-k,1\}}^{T-k}\varepsilon_t^2\left(\theta_0\right)\varepsilon_{t+k}^2\left(\theta_0\right)\right)\\ =\quad&\sum_{k=u}^{T}\frac{1}{k^2}\sigma^2\left(1-\lambda\right)\to 0.\end{aligned}$$

Thus, the result follows from showing that

$$\sum_{k=u}^{T}\frac{1}{k^2}\left(\frac{1}{T^{1/2}}\sum_{t=\max\{[\lambda T]+1-k,1\}}^{T-k}\varepsilon_t\left(\theta_0\right)\varepsilon_{t+k}\left(\theta_0\right)\right)^2$$

is tight in λ. For this, we need to show that

$$\sum_{k=u}^{T}\frac{1}{k^2}\left(\frac{1}{T^{1/2}}\sum_{t=[\lambda_1 T]-k+1}^{[\lambda_2 T]-k}\varepsilon_t\left(\theta_0\right)\varepsilon_{t+k}\left(\theta_0\right)\right)^4\leq|\lambda_2-\lambda_1|^2$$

In particular, $T^{-2}E\sum_{t=[\lambda_1 T]-k}^{[\lambda_2 T]-k}\varepsilon_t^4\left(\theta_0\right)\varepsilon_{t+k}^4\left(\theta_0\right) = T^{-1}T^{-1}\sum_{t=[\lambda_1 T]-k+1}^{[\lambda_2 T]-k}\kappa_4^2 = |\lambda_2-\lambda_1|^2$ and

$$T^{-1}E\sum_{t=[\lambda_1 T]-k+1}^{[\lambda_2 T]-k}\varepsilon_t^2\left(\theta_0\right)\varepsilon_{t+k}^2\left(\theta_0\right)T^{-1}\sum_{s=[\lambda_1 T]-k+1}^{[\lambda_2 T]-k}\varepsilon_s^2\left(\theta_0\right)\varepsilon_{s+k}^2\left(\theta_0\right)=|\lambda_2-\lambda_1|^2.$$

The second term of (2.28) is $o_p(1)$ from the Cauchy Schwarz inequality $[E(XY)\leq\left(EX^2EY^2\right)^{1/2}]$ and from the result about the first term of (2.28).

Next, in direction of ν, given λ,

$$\begin{aligned}
\widetilde{LM}_{\nu,T} &= \left.\frac{\partial}{\partial\nu}L\left(\theta,d,\nu,\mu,\sigma^2,\lambda\right)\right|_{\theta=0,d=\tilde{d}_T,\nu=0,\mu=\tilde{\mu}_T,\sigma^2=\tilde{\sigma}_T^2}\\
&= -\frac{1}{\tilde{\sigma}_T^2}\sum_{t=1}^{T}\left(\Delta_t^{\tilde{d}_T}1D_t(\lambda)\right)\Delta_t^{\tilde{d}_T}\left(y_t-\tilde{\mu}_T\right)\\
&= -\frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)\hat{\varepsilon}_t+\frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\left(\tilde{d}_T-d_0\right)\left(\dot{\Delta}_t^{d}1\right)\tilde{\varepsilon}_t\\
&= -\frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)\tilde{\varepsilon}_t+o_p(T^{-1/2})
\end{aligned}$$

where $\tilde{\varepsilon}_t=\Delta_t^{\tilde{d}_T}\left(y_t-\tilde{\mu}_T\right)$. We use a Taylor expansion around $(\tilde{d}_T,\tilde{\mu}_T)=(d_0,\mu_0)$

$$\begin{aligned}
\tilde{\varepsilon}_t &= \Delta_t^{d_0}\left(y_t-\mu_0\right)+(\tilde{d}_T-d_0)\dot{\Delta}_t^{d_0}\left(y_t-\mu_0\right)+\left(\tilde{\mu}_T-\mu_0\right)\Delta_t^{d_0}1+o_p(T^{-1/2})\\
&= \varepsilon_t\left(\theta_0,\nu_0\right)+\left(\tilde{d}_T-d_0\right)\log\Delta\varepsilon_t\left(\theta_0,\nu_0\right)+\left(\Delta_t^{d_0}1\right)\left(\tilde{\mu}_T-\mu_0\right)\Delta_t^{d_0}1+o_p(T^{-1/2})
\end{aligned}$$

and obtain

$$\begin{aligned}
\widetilde{LM}_{\nu,T} &= -\frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)\varepsilon_t\left(\theta_0,\nu_0\right) && (2.29)\\
&\quad -\frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)\left(\tilde{d}_T-d_0\right)\log\Delta\varepsilon_t\left(\theta_0,\nu_0\right) && (2.30)\\
&\quad +\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)\left(\tilde{\mu}_T-\mu_0\right)\Delta_t^{d_0}1 && (2.31)\\
&\quad +o_p(T^{-1/2}).
\end{aligned}$$

Recalling from Chapter 1 that

$$\begin{aligned}
(\tilde{d}_T-d_0) &= O_p(T^{-1/2})\\
(\tilde{\mu}_T-\mu_0) &= O_p(T^{-1/2+d_0})
\end{aligned}$$

and

$$\dot{\Delta}_t^{d_0}\left(y_t-\mu_0\right)=\log\Delta\Delta_t^{d_0}\left(y_t-\mu_0\right)=\sum_{j=1}^{t}j^{-1}\varepsilon_{t-j}\left(\theta_0,\nu_0\right).$$

The term (2.30) is of smaller order since from (2.20),

$$\begin{aligned}
& E\left(T^{d_0-1/2}\frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}(\Delta_t^{\tilde{d}_T}1)(\tilde{d}_T-d_0)\log\Delta\varepsilon_t\left(\theta_0,\nu_0\right)\right)^2\\
\simeq\; & T^{2d_0-1}E\left(\sum_{t=[\lambda T]+1}^{T}t^{-d}\sum_{j=1}^{t}j^{-1}\left[\left(1-\delta/T^{1/2}D_t\left(\lambda\right)\log\Delta\right)\varepsilon_{t-j}-\eta/T^{1/2-d_0}D_{t-j}\left(\lambda\right)\Delta_{t-j}^{d_0}1\right]\right)^2\\
& \xrightarrow{p} 0.
\end{aligned}$$

Next, from the term (2.31),

$$T^{d_0-1/2}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)^2\left(\tilde{\mu}_T-\mu_0\right)=\frac{\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)^2}{\sum_{t=1}^{T}\left(\Delta_t^{d_0}1\right)^2}T^{d_0-1/2}\sum_{t=1}^{T}\left(\Delta_t^{d_0}1\right)\varepsilon_t\left(\theta_0,\nu_0\right),$$

the first factor converges uniformly in λ (since deterministic) to $(1-\lambda^{1-2d_0})$. Therefore, combining this term with (2.29),

$$\widetilde{LM}_{\nu,T} = -\frac{1}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)\varepsilon_t\left(\theta_0,\nu_0\right)-\frac{1}{\tilde{\sigma}_T^2}\sum_{t=1}^{T}\left(\Delta_t^{d_0}1\right)\varepsilon_t\left(\theta_0,\nu_0\right)\left(1-\lambda^{1-2d_0}\right) \quad (2.32)$$

$$= -\frac{1}{\tilde{\sigma}_T^2}\left[\lambda^{1-2d_0}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)\varepsilon_t\left(\theta_0,\nu_0\right)-\left(1-\lambda^{1-2d_0}\right)\sum_{t=1}^{[\lambda T]}\left(\Delta_t^{d_0}1\right)\varepsilon_t\left(\theta_0,\nu_0\right)\right] \quad (2.33)$$

Using again (2.20), the terms in (2.33) are independent, thus, the term (2.33) converges weakly to

$$\frac{\lambda^{1-2d_0}\left[\tilde{W}_{1/2-d_0}(1)-\tilde{W}_{1/2-d_0}(\lambda)\right]}{\sqrt{(1-2d_0)\,\Gamma^2(1-d_0)}}-\frac{(1-\lambda^{1-2d_0})\tilde{W}_{1/2-d_0}(\lambda)}{\sqrt{(1-2d_0)\,\Gamma^2(1-d_0)}}$$

$$= \frac{\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}(1)-\tilde{W}_{1/2-d_0}(\lambda)}{\sqrt{(1-2d_0)\,\Gamma^2(1-d_0)}},$$

where we use the following convergence result similar to Marinucci and Robinson (1999),

$$T^{d_0-1/2}\sum_{t=1}^{[\lambda T]}\left(\Delta_t^{d_0}1\right)\varepsilon_t=\sum_{t=1}^{[\lambda T]}\pi_{t-1}\left(d_0-1\right)\varepsilon_t\Longrightarrow\frac{\tilde{W}_{1/2-d_0}(\lambda)}{\sqrt{(1-2d_0)\,\Gamma^2(1-d_0)}},$$

with $\tilde{W}_{1/2-d_0}(\lambda)=\int_0^\lambda s^{-d_0}dB(s)$. The fractional Brownian motion $\tilde{W}_{1/2-d_0}(\lambda)$ has the same marginal distribution as the standard one $W_{1/2-d_0}(\lambda)=\int_0^\lambda(\lambda-s)^{-d_0}dB(s)$ but has a different covariance (see the discussion in Chapter 1). In particular,

$$Cov\left(\tilde{W}_{1/2-d_0}(1),\tilde{W}_{1/2-d_0}(\lambda)\right)=\lim_{T\to\infty}T^{2d_0-1}\sum_{t=1}^{[\lambda T]}\left(\Delta_t^{d_0}1\right)^2=\frac{\lambda^{1-2d_0}}{\Gamma(1-d_0)(1-2d_0)} \quad (2.34)$$

rather than

$$Cov\left(W_{1/2-d_0}(1),W_{1/2-d_0}(\lambda)\right)=\frac{1+\lambda^{1-2d_0}}{\Gamma(1-d_0)(1-2d_0)}-E\left[W_{1/2-d_0}(1)-W_{1/2-d_0}(\lambda)\right]^2.$$

Thus $\tilde{W}_{1/2-d_0}(1)-\tilde{W}_{1/2-d_0}(\lambda)$ and $\tilde{W}_{1/2-d_0}(\lambda)$ are independent.

Next, for the drift component, similarly to before, (2.29)

$$\frac{1}{\tilde{\sigma}_T^2}\eta T^{2d_0-1}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)^2\xrightarrow{p}\eta\frac{\left(1-\max\{\lambda,\lambda_0\}^{1-2d_0}\right)}{(1-2d_0)\,\Gamma^2(1-d_0)}$$

and (2.31) converges to

$$\eta\frac{\left(1-\lambda^{1-2d_0}\right)\left(1-\lambda_0^{1-2d_0}\right)}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}$$

both uniformly in λ since they is deterministic. Hence,

$$\begin{aligned} T^{d_0-1/2}\widetilde{LM}_{\nu,T}\left(\lambda\right) &\Longrightarrow \frac{\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}(1)-\tilde{W}_{1/2-d_0}(\lambda)}{\sqrt{(1-2d_0)\Gamma^2(1-d_0)}} \\ &-\eta\frac{\lambda^{1-2d_0}\left(1-\lambda_0^{1-2d_0}\right)-\left(\lambda^{1-2d_0}-\lambda_0^{1-2d_0}\right)_+}{(1-2d_0)\Gamma^2(1-d_0)}. \end{aligned} \tag{2.35}$$

Finally, the standardized test, assuming the memory d is known or consistently estimated,

$$\begin{aligned} & T^{d_0-1/2}\left(\frac{\lambda^{1-2d_0}(1-\lambda^{1-2d_0})}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}\right)^{-1/2}\widetilde{LM}_{\nu,T}\left(\lambda\right) \\ \Longrightarrow\ & \frac{\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}\left(1\right)-\tilde{W}_{1/2-d_0}\left(\lambda\right)}{\left(\lambda^{1-2d_0}(1-\lambda^{1-2d_0})\right)^{1/2}}-\eta\left(\frac{\lambda^{1-2d_0}\left(1-\lambda_0^{1-2d_0}\right)-\left(\lambda^{1-2d_0}-\lambda_0^{1-2d_0}\right)_+}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}\right)^{1/2} \end{aligned}$$

Finally, combining the two dimensions, the LM test for a break in both parameters, (2.25) and (2.35) and defining

$$D_T=diag\left(T^{-1/2},T^{-1/2+d_0}\right), \tag{2.36}$$

from the results about the derivatives, (2.35) and (2.25), we obtain

$$\begin{aligned} & D_T\widetilde{LM}_T\left(\lambda\right)=D_T\frac{\delta L}{\delta\psi}\left(\lambda\right)|_{\theta=0,\nu=0} \\ \Longrightarrow\ & \left(\begin{array}{c} \frac{\pi}{\sqrt{6}}\left(\lambda B\left(1\right)-B\left(\lambda\right)\right)-\delta\left(\lambda\left(1-\lambda_0\right)-\left(\lambda-\lambda_0\right)_+\right)\frac{\pi^2}{6} \\ \frac{\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}(1)-\tilde{W}_{1/2-d_0}(\lambda)}{\sqrt{(1-2d_0)\Gamma^2(1-d_0)}}-\eta\frac{\lambda^{1-2d_0}\left(1-\lambda_0^{1-2d_0}\right)-\left(\lambda^{1-2d_0}-\lambda_0^{1-2d_0}\right)_+}{(1-2d_0)\Gamma^2(1-d_0)} \end{array}\right), \end{aligned}$$

with the two components being uncorrelated. In addition, the Hessian uniformly in λ

$$D_T\frac{\delta^2 L}{\delta\psi^2}D_T\xrightarrow{p}-\left(\begin{array}{cc} \lambda\left(1-\lambda\right)\frac{\pi^2}{6} & 0 \\ 0 & \frac{\lambda^{1-2d_0}\left(1-\lambda^{1-2d_0}\right)}{(1-2d_0)\Gamma^2(1-d_0)} \end{array}\right).$$

Hence,

$$LM_T(\lambda) \implies \frac{\left(\lambda B(1) - B(\lambda) - \delta\left(\lambda(1-\lambda_0) - (\lambda-\lambda_0)_+\right)\frac{\pi}{\sqrt{6}}\right)^2}{\lambda(1-\lambda)}$$
$$+\frac{\left(\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}(1) - \tilde{W}_{1/2-d_0}(\lambda) - \eta\frac{\lambda^{1-2d_0}\left(1-\lambda_0^{1-2d_0}\right)-\left(\lambda^{1-2d_0}-\lambda_0^{1-2d_0}\right)_+}{\sqrt{1-2d_0}\Gamma(1-d_0)}\right)^2}{\lambda^{1-2d_0}\left(1-\lambda^{1-2d_0}\right)}$$

Proof of Corollary 23

As a special case of Theorem 22, for a known break fraction λ, the distribution of the derivative in the direction of the memory is

$$T^{-1/2}LM_{d,T} \xrightarrow{d} N\left(\delta\lambda_0(1-\lambda_0)\frac{\pi^2}{6}, \lambda_0(1-\lambda_0)\frac{\pi^2}{6}\right)$$

since

$$Var\left[\lambda B(1) - B(\lambda)\right] = \lambda^2 + \lambda - 2\lambda^2,$$

where the drift and the AVar is symmetric around $\lambda = \frac{1}{2}$. Further, the distribution of the derivative in the direction of the level is

$$T^{d_0-1/2}LM_{\nu,T} \xrightarrow{d} N\left(\eta\frac{\lambda_0^{1-2d_0}(1-\lambda_0^{1-2d_0})}{(1-2d_0)\Gamma^2(1-d_0)}, \frac{(\lambda_0^{1-2d_0})(1-\lambda_0^{1-2d_0})}{(1-2d_0)\Gamma^2(1-d_0)}\right),$$

since

$$Var\left[\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}(1) - \tilde{W}_{1/2-d_0}(\lambda)\right] = \left(\lambda^{1-2d_0}\right)^2 + \lambda^{1-2d_0} - 2\left(\lambda^{1-2d_0}\right)^2$$

because of the particular covariance structure of $\tilde{W}_{1/2-d_0}(.)$, in (2.34).

Proof of Proposition 24

Next, we analyze the case that the parameters in the first regime (d_0, μ_0) under the H_0 are known rather than estimated.

Part a)

Starting with an unknown break fraction, if d_0 was known, we would have no estimation effect and the second term in (2.23) would drop. For the derivative in the direction of θ,

$$T^{-1/2}LM_{d,T}(\lambda) \implies -\delta(1-\max\{\lambda_0,\lambda\})\frac{\pi^2}{6} + \frac{\pi}{\sqrt{6}}(B(1) - B(\lambda))$$

rather than (2.25). In this case, drift and AVar are both proportional to the size of the sample that is used to test $\theta \neq 0$, $t = [\lambda T]+1,\dots,T$. The standardized test statistic in this case would be then

$$
\begin{aligned}
& T^{-1/2}\left(\left(1-\lambda\right)\frac{\pi^2}{6}\right)^{-1/2} LM_{d,T} \\
\Longrightarrow \quad & -\delta\left(\left(1-\max\{\lambda_0,\lambda\}\right)\frac{\pi^2}{6}\right)^{1/2}+\frac{\left(B\left(1\right)-B\left(\lambda\right)\right)}{\left(1-\lambda\right)^{1/2}}.
\end{aligned}
$$

On the other hand if μ_0 was known, we would have no estimation effect for μ_0 and the second term in expression (2.32) would drop. For the derivative in the direction of ν,

$$
T^{-1/2+d_0}LM_{\nu,T}\left(\lambda\right)\Longrightarrow -\eta\frac{1-\max\{\lambda_0,\lambda\}^{1-2d_0}}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}+\frac{\tilde{W}_{1/2-d_0}\left(1\right)-\tilde{W}_{1/2-d_0}\left(\lambda\right)}{\sqrt{1-2d_0}\Gamma\left(1-d_0\right)}
$$

rather than (2.35). Combining both leads to the desired result.

b) For a known break fraction, note that the derivatives of the likelihood function are Gaussian and

$$
Var\left(T^{d_0-1/2}LM_{\theta T}\right)=\left(1-\lambda_0\right)\frac{\pi^2}{6}
$$

and

$$
Var(T^{d_0-1/2}LM_{\nu T})=\frac{1-\max\{\lambda_0,\lambda\}^{1-2d_0}}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}
$$

. Therefore, if both parameters are known

$$
LM_T\xrightarrow{d}\chi_2^2\left(c\right)
$$

with a noncentrality parameter

$$
c=\delta^2\left(1-\lambda_0\right)\frac{\pi^2}{6}+\eta^2\frac{1-{\lambda_0}^{1-2d_0}}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}.
$$

Proof of Proposition 25

Part a)

Consider the DGP (2.8). Then, we find from a similar argument as in Chapter 1 for (2.25), under H_{1T},

$$
T^{-1/2}\widetilde{LM}_{d,T}\Longrightarrow\delta\left(\lambda\left(1-\lambda_0\right)-\left(\lambda-\lambda_0\right)_+\right)\omega^2+\omega\left(\lambda B\left(1\right)-B\left(\lambda\right)\right),
$$

where $\omega^2=\frac{\pi^2}{6}-\kappa'\Phi\kappa$ is defined as in Lobato and Velasco (2006). Since the estimation of the level is independent from the short memory dynamics, the derivative (2.35) in the direction of ν is unaffected. Combining both derivatives leads to the result.

Part b)

Noticing that $T^{-1/2}\widetilde{LM}_{d,T}$ is Gaussian with a variance $\lambda(1-\lambda)\omega^2$ establishes the result.

Proof of Proposition 26

Part a)

For an unknown break fraction, the should be

$$\left(\frac{\partial L(\theta, d, \mu, \nu, \sigma^2, \lambda)}{\partial d}, \frac{\partial L(\theta, d, \mu, \nu, \sigma^2, \lambda)}{\partial \boldsymbol{\beta}}\right)' \Rightarrow \Xi\left[\lambda B_{p+1}(1) - B_{p+1}(\lambda)\right].$$

From a similar argument as the one in Theorem 22, the local drift amounts to

$$\left(\lambda(1-\lambda_0) - (\lambda-\lambda_0)_+\right)\begin{pmatrix} \delta & \gamma' \end{pmatrix}\Xi\begin{pmatrix} \delta \\ \gamma \end{pmatrix}^{1/2}$$

Part b)

For a known break fraction λ_0, the test converges to

$$LM_T \xrightarrow{d} \chi^2_{2+p}(c),$$

with

$$c = \lambda_0(1-\lambda_0)\begin{pmatrix} \delta & \gamma' \end{pmatrix}\Xi\begin{pmatrix} \delta \\ \gamma \end{pmatrix} + \eta^2\frac{\lambda_0^{1-2d_0}\left(1-\lambda_0^{1-2d_0}\right)}{(1-2d_0)\Gamma^2(1-d_0)}.$$

Proof of Proposition 27

Part a)

We proof the consistency for the tests under H_1, (2.1), of breaks in both parameters. The proof works by showing that the rate needed for convergence of the numerator of the test statistic is the square of the one of the test statistic. Thus, the test statistic diverges at rates T and T^{1-2d_0} respectively. First, under H_1, the estimators $\tilde{\mu}_T$ and $\tilde{d}_T$ converge to some weighted averages

$$\tilde{\mu} \xrightarrow{p} \bar{\mu} = \lambda_0^{1-2d_0}\mu_0 + (1-\lambda_0^{1-2d_0})\mu_1 \tag{2.37}$$

and (2.11) respectively.

For $LM_{\theta,T}(\lambda)$, the term (2.31),

$$T^{1/2}\sum_{k=1}^{T-1}\left\{\frac{1}{k}\tilde{\rho}_k^*(\varepsilon_t(\theta_0)) + \frac{1}{k^2}(1-\lambda)\left(\tilde{d}_T - d_0\right)\right\}$$

is of order $T^{1/2}$ since $(\tilde{d}_T - d_0) = O(1)$. Combining this with the behavior of the denominator, the test statistic LM (2.4) for $\vartheta = \mu$ and d, μ diverges at rate T.

For testing for a break in the level, we distinguish between the cases of a breaking memory (case a)) and not breaking memory (case c)). For the latter, the memory is still consistently estimated, $T^{1/2}(\tilde{d}_T - d_0) = O_p(1)$ and for $LM_{\nu,T}(\lambda)$, the term (2.31),

$$T^{d_0-1/2} \sum_{t=[\lambda T]+1}^{T} \left(\Delta_t^{d_0} 1\right)^2 (\tilde{\mu}_T - \mu_0) = T^{2d_0-1} \sum_{t=[\lambda T]+1}^{T} \left(\Delta_t^{d_0} 1\right)^2 T^{1/2-d_0} (\tilde{\mu}_T - \mu_0) = O_p\left(T^{1/2-d_0}\right).$$

Combining it with the rate of convergence of the denominator, the test statistic LM (2.4) for $j = \mu$ diverges at rate T^{1-2d_0}. On the other hand, if memory is also breaking, the memory is not consistently estimated anymore. However, this does not affect the consistency of the level estimation.

Proof of Theorem 28

Part a)

First consider an unknown break fraction under the local alternative H_{1T}. The break fraction is estimated by

$$\hat{\lambda} = \arg\max_{\lambda} W_T(\lambda),$$

using the observations of the second regime. Recall that

$$y_t - \mu_0 = \nu D_t(\lambda) + \Delta_t^{-d_0-\theta D_t(\lambda)} \varepsilon_t = \eta / T^{1/2-d_0} D_t(\lambda) + \Delta_t^{-d_0+\delta/\sqrt{T} D_t(\lambda)} \varepsilon_t$$

and thus

$$\Delta_t^{d_0}(y_t - \mu_0) = \eta / T^{1/2-d_0} \Delta_t^{d_0} 1 D_t(\lambda) + \Delta_t^{\delta/\sqrt{T} D_t(\lambda)} \varepsilon_t.$$

Next, write

$$\hat{\vartheta}_1 = \frac{\sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^{\theta}}{\theta}\right] \left(\Delta_t^{\delta/\sqrt{T}} \varepsilon_t + \eta / T^{1/2-d_0} \Delta_t^{d_0} 1\right)\right) \left(\Delta_t^{\delta/\sqrt{T}} \varepsilon_t + \eta / T^{1/2-d_0} \Delta_t^{d_0} 1\right)}{\sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^{\theta}}{\theta}\right] \left(\Delta_t^{\delta/\sqrt{T}} \varepsilon_t + \eta / T^{1/2-d_0} \Delta_t^{d_0} 1\right)\right)^2} + o_p\left(T^{-1/2}\right).$$

First, we can show for the denominator

$$\frac{1}{T} \sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^{\theta}}{\theta}\right] \left(\Delta_t^{-\theta} \varepsilon_t + \Delta_t^{d_0} \nu\right)\right)^2 \xrightarrow{p} (1-\lambda) \sum_{j=1}^{\infty} \pi_j^2(\theta),$$

uniformly in λ. Next, for the numerator from a functional central limit theorem for martingale difference sequences,

$$T^{-1/2} \sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^{\theta}}{\theta}\right] \varepsilon_t\right) \varepsilon_t \Longrightarrow \sqrt{\sum_{j=1}^{\infty} \pi_j^2(\theta)} \left[B(1) - B(\lambda)\right].$$

From Lobato and Velasco (2007), the local drift term adds a term

$$-\delta K\left(d_0+\theta\right)(1-\lambda),$$

where

$$K\left(d_0+\theta\right)=\sum_{i=1}^{\infty}\frac{\pi_i\left(\theta\right)}{i\left(\theta\right)}.$$

Thus, the local drift for $\theta\to 0$,

$$-\delta\left(1-\lambda\right)^{1/2}\frac{\pi}{\sqrt{6}}.$$

Next, define

$$h\left(d_0+\theta\right)=\frac{\sum_{i=1}^{\infty}i^{-1}\pi_i\left(\theta\right)}{\sqrt{\sum_{i=1}^{\infty}\pi_i^2\left(\theta\right)}},$$

which for $\theta\to 0$,

$$h\left(d_0\right)=\sqrt{\pi^2/6}.$$

Thus,

$$T^{1/2}\hat{\vartheta}_1\Longrightarrow -\delta K\left(d_0+\theta\right)(1-\lambda)+\frac{\left[B\left(1\right)-B\left(\lambda\right)\right]}{\left(1-\lambda\right)\sqrt{\sum_{j=1}^{\infty}\pi_j^2\left(\theta\right)}}$$

and finally for $\theta\to 0$ the resulting t-statistic

$$t_1\Longrightarrow -\frac{\delta\left(\lambda\left(1-\lambda_0\right)-\left(\lambda-\lambda_0\right)_+\right)}{\left(1-\lambda\right)^{1/2}}\frac{\pi}{\sqrt{6}}+\frac{\left[B\left(1\right)-B\left(\lambda\right)\right]}{\left(1-\lambda\right)^{1/2}}.$$

Thus, the test for only a break in the memory,

$$t_1^2\Longrightarrow\frac{\left[B\left(1\right)-B\left(\lambda\right)-\delta\left(\left(1-\lambda_0\right)-\left(\lambda-\lambda_0\right)_+\right)\frac{\pi}{\sqrt{6}}\right]^2}{\left(1-\lambda\right)}.$$

Next, we consider the estimator ϑ_2,

$$T^{1/2-d_0}\hat{\vartheta}_2=\frac{T^{d_0-1/2}\sum\limits_{t=[\lambda T]+1}^{T}\Delta_t^{d_0}1\left(\eta/T^{1/2-d_0}\Delta_t^{d_0}1+\Delta_t^{\delta/\sqrt{T}}\varepsilon_t\right)}{T^{2d_0-1}\sum\limits_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)^2}+o_p\left(1\right).$$

The denominator converges uniformly in λ (since deterministic) to

$$\frac{1-\lambda^{1-2d_0}}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}.$$

For the numerator, from Marinucci and Robinson (1999),

$$
\begin{aligned}
& T^{d_0-1/2} \sum_{t=[\lambda T]+1}^{T} \Delta_t^{d_0} 1 \left(\eta / T^{1/2-d_0} \Delta_t^{d_0} 1 + \Delta_t^{\delta/\sqrt{T}} \varepsilon_t \right) \\
= \; & T^{d_0-1/2} \sum_{t=[\lambda T]+1}^{T} \Delta_t^{d_0} 1 \left(\eta / T^{1/2-d_0} \Delta_t^{d_0} 1 \right) + T^{d_0-1/2} \sum_{t=\lambda T}^{T} \Delta_t^{d_0} 1 \varepsilon_t \\
\Longrightarrow \; & \frac{1}{(1-2d_0)\Gamma^2(1-d_0)} + \frac{\tilde{W}_{1/2-d_0}(1) - \tilde{W}_{1/2-d_0}(\lambda)}{\sqrt{(1-2d_0)\Gamma^2(1-d_0)}},
\end{aligned}
$$

with $\tilde{W}_{1/2-d_0}(\lambda) = \int_0^\lambda s^{-d_0} dB(s)$ discussed in the Proof of Theorem 1. Combining the numerator and the denominator, we obtain

$$
\begin{aligned}
T^{1/2}\hat{\vartheta}_2 \;\Longrightarrow\; & -\eta \frac{(1-\lambda_0^{1-2d_0}) - (\lambda^{1-2d_0} - \lambda_0^{1-2d_0})_+}{\sqrt{1-2d_0}\Gamma(1-d_0)(1-\lambda^{1-2d_0})} \\
& + \frac{\sqrt{(1-2d_0)\Gamma^2(1-d_0)}\left[\tilde{W}_{1/2-d_0}(1) - \tilde{W}_{1/2-d_0}(\lambda)\right]}{(1-\lambda^{1-2d_0})}.
\end{aligned}
$$

Thus, the test for a break only in the level behaves as

$$
t_2^2 \Longrightarrow \frac{\left[\tilde{W}_{1/2-d_0}(1) - \tilde{W}_{1/2-d_0}(\lambda) - \eta \frac{(1-\lambda_0^{1-2d_0})-(\lambda^{1-2d_0}-\lambda_0^{1-2d_0})_+}{\sqrt{1-2d_0}\Gamma(1-d_0)}\right]^2}{(1-\lambda^{1-2d_0})}.
$$

Finally, combining t_1^2 and t_2^2 leads to the stated result.

Part b)

The proof follows from Part a) in a similar manner as in Theorem 22.

Proof of Theorem 29

First, note that an unknown $\theta = d_2 - d_0$ can be dealt with exactly as in Lobato and Velasco (2007). Equally, an estimated ν does not affect the asymptotic distribution.

In the case of an unknown d_0 and μ_0, we would have to estimate them by the CSS estimator using observations from the first regime. For simplicity, we analyze the test statistic under the H_0, the case under the local alternative behaves accordingly. First, we start again with the estimator ϑ_1,

$$
T^{1/2}\hat{\vartheta}_1 = \frac{T^{-1/2} \sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^\theta}{\theta} \right] \Delta_t^{\hat{d}_0} (y_t - \hat{\mu}_0) \right) \Delta_t^{\hat{d}_0} (y_t - \hat{\mu}_0)}{\frac{1}{T} \sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^\theta}{\theta} \right] \Delta_t^{\hat{d}_0} (y_t - \hat{\mu}_0) \right)^2} + o_p(1). \tag{2.38}
$$

The numerator

$$
\begin{aligned}
& T^{-1/2} \sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^{\theta}}{\theta} \right] \Delta_t^{\hat{d}_0} \left(y_t - \hat{\mu}_0 \right) \right) \Delta_t^{\hat{d}_0} y_t \\
= \; & T^{-1/2} \sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^{\theta}}{\theta} \right] \Delta_t^{\hat{d}_0} \left(y_t - \hat{\mu}_0 \right) \right) \left[\varepsilon_t + \sum_{j=1}^{t-1} \pi_j \left(\hat{d}_0 - d_0 \right) \varepsilon_{t-j} \right] \quad (2.39)
\end{aligned}
$$

consists of two uncorrelated terms. The first one corresponds to the term when d_0 known. For the second term, from a Taylor approximation

$$
\begin{aligned}
& T^{-1/2} \sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^{\theta}}{\theta} \right] \Delta_t^{\hat{d}_0} \left(y_t - \hat{\mu}_0 \right) \right) \sum_{j=1}^{t-1} \pi_j \left(\hat{d}_0 - d_0 \right) \varepsilon_{t-j} \\
= \; & T^{1/2} \left(\hat{d}_0 - d_0 \right) T^{-1} \sum_{t=[\lambda T]+1}^{T} \left(\left[\frac{1-\Delta_t^{\theta}}{\theta} \right] \Delta_t^{\hat{d}_0} \left(y_t - \hat{\mu}_0 \right) \right) \sum_{j=1}^{t-1} \dot{\pi}_j \left(\dot{\theta} \right) \varepsilon_{t-j},
\end{aligned}
$$

where $0 < |\dot{\theta}| < |\hat{d} - d_0|$. The first factor converges from (2.14) to

$$
T^{1/2} \left(\hat{d}_0 - d_0 \right) \Longrightarrow \frac{\sqrt{6}}{\pi} \frac{B(\lambda)}{\lambda}.
$$

The second factor converges by a LLN to

$$
\sum_{j=1}^{\infty} \frac{1}{\theta} \pi_j(\theta) \pi_j(\dot{\theta}),
$$

which behaves for $\theta \longrightarrow 0$, and $\hat{d}_0 \longrightarrow d_0$,

$$
\sum_{j=1}^{\infty} \frac{1}{\theta} \pi_j(\theta) \pi_j(\dot{\theta}) \xrightarrow{p} -\frac{\pi^2}{6}.
$$

Finally, from combining the first and second terms in (2.39) with the denominator of (2.38),

$$
T^{1/2} \hat{\vartheta}_1 \Longrightarrow \frac{\sqrt{6}}{\pi} \frac{\lambda B(1) - B(\lambda)}{\lambda(1-\lambda)}.
$$

This leads to a test statistic under unknown (d_0, μ_0)

$$
\begin{aligned}
W_{1T} \quad & \Longrightarrow \quad \frac{B(1) - B(\lambda)}{1-\lambda} - \frac{B(\lambda)}{\lambda} = \frac{\lambda B(1) - \lambda B(\lambda) - B(\lambda) + \lambda B(\lambda)}{\lambda(1-\lambda)} \\
& = \quad \frac{\lambda B(1) - B(\lambda)}{\lambda(1-\lambda)}. \quad (2.40)
\end{aligned}
$$

Next, we consider the estimator ϑ_2, where we again neglect the effect of estimating d_0,

$$\begin{aligned} T^{1/2-d_0}\hat{\vartheta}_2 &= \frac{T^{d_0-1/2}\sum\limits_{t=[\lambda T]+1}^{T}\Delta_t^{d_0}1\Delta_t^{d_0}\left(y_t-\hat{\mu}_0\right)}{T^{2d_0-1}\sum\limits_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)^2} \\ &= \frac{T^{d_0-1/2}\sum\limits_{t=[\lambda T]+1}^{T}\Delta_t^{d_0}1\Delta_t^{d_0}\left(y_t-\mu_0\right)}{T^{2d_0-1}\sum\limits_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)^2} - T^{1/2-d_0}\left(\hat{\mu}_0-\mu_0\right). \end{aligned}$$

The first term behaves as described in Theorem 28, the second term behaves as described in expression (2.14). Putting the two terms together, we get to

$$\begin{aligned} & \frac{\sigma\Gamma\left(1-d_0\right)\sqrt{(1-2d_0)}\left[\tilde{W}_{1/2-d_0}\left(1\right)-\tilde{W}_{1/2-d_0}\left(\lambda\right)\right]}{(1-\lambda^{1-2d_0})} - \frac{\sigma\Gamma\left(1-d_0\right)\sqrt{(1-2d_0)}\tilde{W}_{1/2-d_0}\left(\lambda\right)}{\lambda^{1-2d_0}} \\ = \; & \sigma\Gamma\left(1-d_0\right)\sqrt{(1-2d_0)}\frac{\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}\left(1\right)-\tilde{W}_{1/2-d_0}\left(\lambda\right)}{(1-\lambda^{1-2d_0})\lambda^{1-2d_0}}. \end{aligned}$$

This leads to a test statistic

$$W_{2T} \Longrightarrow \frac{\lambda^{1-2d_0}\tilde{W}_{1/2-d_0}\left(1\right)-\tilde{W}_{1/2-d_0}\left(\lambda\right)}{(1-\lambda^{1-2d_0})\lambda^{1-2d_0}}. \tag{2.41}$$

Finally, since the estimation is asymptotically uncorrelated, the test for a break in both parameters corresponds to the sum of (2.40) and (2.41).

Proof of Proposition 30

The proof for an equivalent two-step procedure follows from combining the Appendix 2 of Lobato and Velasco (2007) and our Theorem 28. The equivalence of the one step procedure follows from arguments in Dolado et al (2007).

Proof of Proposition 31

First, under the fixed alternative of one break in the memory in any direction, $\theta \neq 0$, for $t = [\lambda T]+1, ..., T$,

$$\Delta_t^{d_0}\left(y_t-\mu_0\right) = \Delta_t^{d_0}\nu + \Delta_t^{-\theta}\varepsilon_t. \tag{2.42}$$

Next, we look again at the estimator ϑ_1,

$$\begin{aligned}\hat{\vartheta}_1 &= \frac{\frac{1}{T}\sum_{t=[\lambda T]+1}^{T}\left(\left[\frac{1-\Delta_t^{\theta}}{\theta}\right]\Delta_t^{d_0}\left(y_t-\mu_0\right)\right)\Delta_t^{d_0}\left(y_t-\mu_0\right)}{\frac{1}{T}\sum_{t=[\lambda T]+1}^{T}\left(\left[\frac{1-\Delta_t^{\theta}}{\theta}\right]\Delta_t^{d_0}\left(y_t-\mu_0\right)\right)^2} o_p\left(T^{1/2-d_0}\right)\\ &= \frac{\frac{1}{T}\sum_{t=[\lambda T]+1}^{T}\left(\frac{1}{\theta}\sum_{j=1}^{t-1}\pi_j\left(\theta\right)\varepsilon_t\right)\left(\sum_{j=0}^{t-1}\pi_j\left(\theta\right)\varepsilon_{t-j}\right)}{\frac{1}{T}\sum_{t=[\lambda T]+1}^{T}\left(\frac{1}{\theta}\sum_{j=1}^{t-1}\pi_j\left(\theta\right)\varepsilon_t\right)^2}+o_p\left(T^{1/2-d_0}\right),\end{aligned}$$

where the equality comes from substituting (2.42). The denominator converges again from a LLN to $(1-\lambda)\sum_{j=1}^{\infty}\pi_j^2\left(\theta\right)$ and the numerator converges by mean square convergence to $\theta\sum_{j=1}^{t-1}\pi_j^2\left(\theta\right)$. Further, from the consistency of the estimator $\hat{d}_1$, $|\theta|=|\hat{d}_1-d_0|>0$. Thus, $\hat{\vartheta}_1=O_p\left(1\right)$ and

$$\begin{aligned}t_1 &= O_p(T^{1/2}) \text{ and}\\ W_{1,T} &= O_p\left(T\right).\end{aligned}$$

In consequence, a two sided t-test and the Wald test are consistent for breaks in both directions.

Next, we consider the test for a break in the level. Under the alternative, we have for $t=[\lambda T]+1,...,T$,

$$\Delta_t^{d_0}\left(y_t-\mu_0\right)=\varepsilon_t+\nu\Delta_t^{d_0}1$$

and

$$\vartheta_2=\frac{\sum_{t=[\lambda T]+1}^{T}\Delta_t^{d_0}1\left(\varepsilon_t+\Delta_t^{d_0}\nu\right)}{\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{d_0}1\right)^2}=O_p\left(T^{d_0-1/2}\right)+\nu$$

and

$$\begin{aligned}t_2 &= O_p(T^{1/2-d_0}) \text{ and}\\ W_{2,T} &= O_p\left(T^{1-2d_0}\right).\end{aligned}$$

The test is again consistent for breaks in both directions.

Proof of Proposition 32

Part a)

We start with the LM test in the direction of the memory. Recall that the residuals required to implement the LM test statistic are in (8). Then, for the DGP (2.1), the residuals under the null become,

$$\begin{cases} \Delta_t^{\tilde{d}}\left(y_t-\tilde{\mu}\right)=\Delta_t^{\tilde{d}-d_0}\Delta_t^{d_0}\left(y_t-\mu_0+\mu_0-\tilde{\mu}\right)=\Delta_t^{\tilde{d}-d_0}\varepsilon_t+\left(\mu_0-\tilde{\mu}\right)\Delta_t^{\tilde{d}}1 \\ \Delta_t^{\tilde{d}}\left(y_t-\tilde{\mu}\right)=\Delta_t^{\tilde{d}-d_1}\Delta_t^{d_1}\left(y_t-\mu_0+\mu_0-\tilde{\mu}\right)=\Delta_t^{\tilde{d}-d_1}\varepsilon_t+\left(\mu_1-\tilde{\mu}\right)\Delta_t^{\tilde{d}}1 \end{cases}.$$

As a result, the variance estimator is

$$\tilde{\sigma}_T^2=\frac{1}{T}\sum_{t=1}^{T}\tilde{\varepsilon}_t^2=\frac{1}{T}\sum_{t=1}^{[\lambda T]}\left(\Delta_t^{\tilde{d}-d_0}\varepsilon_t+\left(\mu_0-\tilde{\mu}\right)\Delta_t^{\tilde{d}}1\right)^2+\frac{1}{T}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{\tilde{d}-d_1}\varepsilon_t+\left(\mu_1-\tilde{\mu}\right)\Delta_t^{\tilde{d}}1\right)^2.$$

If the memory is not breaking, the variance estimator $\tilde{\sigma}_T^2$ converges to σ^2 and if the memory breaks,

$$\begin{aligned} \tilde{\sigma}_T^2 &= \frac{1}{T}\sum_{t=1}^{[\lambda T]}\left(\Delta_t^{\tilde{d}-d_0}\varepsilon_t\right)^2+\frac{1}{T}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{\tilde{d}-d_1}\varepsilon_t\right)^2+o_p\left(1\right) \\ &\xrightarrow{p} \lambda\sum_{j=0}^{\infty}\pi_j^2\left(\tilde{d}-d_0\right)\sigma^2+\left(1-\lambda\right)\sum_{j=0}^{\infty}\pi_j^2\left(\tilde{d}-d_1\right)\sigma^2\equiv\left(1+\xi\right)\sigma^2 \end{aligned} \tag{2.43}$$

Further, note that if $d_0=d_1$, the estimator $\tilde{d}$ converges in probability to d_0, while the estimator of the level satisfies (2.37). Moreover, for a break in the memory, we have that, $\tilde{d}\xrightarrow{p}\bar{d}$, where $\bar{d}$ is defined in (2.11).

As regards case (bd), following Lobato and Velasco (2008), we calculate the autocorrelation of $\Delta_t^{\tilde{d}}\left(y_t-\tilde{\mu}\right)$ as

$$\begin{aligned} \tilde{\rho}_k^*\left(\tilde{\varepsilon}_t\right) &= \frac{\tilde{\sigma}_T^{-2}}{T}\sum_{t=\max\{[\lambda T]+1-k,1\}}^{T-k}\tilde{\varepsilon}_t\tilde{\varepsilon}_{t+k} \\ &= \frac{\tilde{\sigma}_T^{-2}}{T}\sum_{t=\max\{[\lambda T]+1-k,1\}}^{T-k}\Delta_t^{\tilde{d}-d_1}\varepsilon_t\Delta_{t+k}^{\tilde{d}-d_1}\varepsilon_{t+k} \\ &\xrightarrow{p}\left(1-\lambda\right)\frac{\Gamma\left(1+\bar{d}-d_1\right)}{\Gamma\left(d_1-\bar{d}\right)}\frac{\Gamma\left(j+d_1-\bar{d}\right)}{\Gamma\left(j+1+\bar{d}-d_1\right)} \end{aligned}$$

and thus,

$$\underset{T\to\infty}{p\lim}\frac{LM}{T}=\left(1-\lambda\right)\frac{6}{\pi^2}\frac{\Gamma^2\left(1+\bar{d}-d_1\right)}{\Gamma^2\left(d_1-\bar{d}\right)}\left(\sum_{j=1}^{\infty}\frac{\Gamma\left(j+d_1-\bar{d}\right)}{j\Gamma\left(j+1+\bar{d}-d_1\right)}\right)^2. \tag{2.44}$$

Part b)

Next, we consider the properties of the Wald test under fixed alternatives. As regards case (bd), in line again with Lobato and Velasco (2008), we can calculate the population correlation coefficient between regressor and regressand as,

$$p\lim_{T\to\infty}\frac{W}{T}=(1-\lambda)\sum_{j=1}^{\infty}\pi_j^2\left(d_0-d_1\right)=(1-\lambda)\left[\frac{\Gamma\left(1+2\left(d_0-d_1\right)\right)}{\Gamma^2\left(1+\left(d_0-d_1\right)\right)}-1\right] \tag{2.45}$$

Part c)

Next, we analyze the behaviour of the LR test, in terms of the Chow-type F-test proposed by Chapter 1. In the notation of Chapter 1,

$$F=\frac{SSR_0-SSR_1\left(\lambda\right)}{\frac{1}{T}SSR_1\left(\lambda\right)},$$

where

$$\frac{1}{T}SSR_k\left(\lambda\right)=\hat{\sigma}^2\xrightarrow{p}\sigma^2$$

and

$$SSR_0-SSR_1\left(\lambda\right)=D^R\left(1,2\right)-D^U\left(1,1\right)-D^U\left(2,2\right).$$

First, for case (bd), we follow Lobato and Velasco (2008) again. to show that,

$$p\lim_{T\to\infty}\frac{F}{T}=(1-\lambda)\log\left(\frac{\Gamma\left(1+2\left(\bar{d}-d_1\right)\right)}{\Gamma^2\left(1+\left(\bar{d}-d_1\right)\right)}\right) \tag{2.46}$$

Proof of Proposition 33

Part a)

With regard to case (bl), the numerator of the LM test statistic becomes,

$$\begin{aligned}\frac{T^{2\tilde{d}_T-1}}{\tilde{\sigma}_T^2}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{\tilde{d}_T}1\right)\hat{\varepsilon}_t &= \left(\frac{\mu_1-\tilde{\mu}}{\tilde{\sigma}_T^2}\right)T^{2\tilde{d}_T-1}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{\tilde{d}_T}1\right)^2\\ &\xrightarrow{p}\left(\frac{\mu_1-\tilde{\mu}}{\sigma^2}\right)\frac{1-\lambda^{1-2d_0}}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}.\end{aligned}$$

Hence,

$$T^{2\tilde{d}_T-1}LM\xrightarrow{p}\left(\mu_1-\tilde{\mu}\right)^2\frac{1-\lambda^{1-2d_0}}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)},$$

which, using (2.37), equals

$$\begin{aligned}&\frac{\left(\mu_1-\left(\lambda^{1-2d_0}\mu_0+(1-\lambda^{1-2d_0})\mu_1\right)\right)^2\left(1-\lambda^{1-2d_0}\right)}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}\\ =\;&\left(\mu_1-\mu_0\right)^2\frac{\lambda_1^{1-2d_0}(1-\lambda^{1-2d_0})}{\left(1-2d_0\right)\Gamma^2\left(1-d_0\right)}.\end{aligned}$$

Part b)

Next, for the case (bl), we obtain the t-statistic of ϑ_2 in the regression:

$$\Delta_t^{d_0}(y_t - \mu_0) = \vartheta_1 \left[\frac{1-\Delta_t^{\theta D_t(\lambda)}}{\theta}\right] \Delta_t^{d_1}(y_t - \mu_1) + \vartheta_2 \Delta_t^{d_1} 1 + \varepsilon_t,$$

leading to the t_{ϑ_2} statistic

$$\begin{aligned} T^{1/2-d_0} t_{\vartheta_2} &= T^{1/2-d_0} \frac{\hat{\vartheta}_2}{\hat{\sigma}_{\vartheta_2}} = \frac{T^{2d_0-1}\sum_{t=1}^{T}(\mu_1-\mu_0)\left(D_t(\lambda)\Delta_t^{d_0}1\right)^2}{\sqrt{T^{2d_0-1}\sum_{[\lambda T]+1}^{T}\left(D_t\Delta_t^{d_0}1\right)^2}} \\ &\xrightarrow{p} \frac{\left(\frac{\mu_1-\mu_0}{\sigma}\right)\frac{1-\lambda^{1-2d_0}}{(1-2d_0)\Gamma^2(1-d_0)}}{\sqrt{\frac{1-\lambda^{1-2d_0}}{(1-2d_0)\Gamma^2(1-d_0)}}} \end{aligned}$$

where, to compute

$$\hat{\sigma}_{\vartheta_2}^2 = \frac{\hat{\sigma}^2}{T^{2d_0-1}\sum_{t=[\lambda T]+1}^{T}\left(D_t\Delta_t^{d_0}1\right)^2},$$

we use the error variance estimator under the alternative, namely,

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{[\lambda T]}\left(\Delta_t^{\hat{d}_0-d_0}\varepsilon_t\right)^2 + \frac{1}{T}\sum_{t=[\lambda T]+1}^{T}\left(\Delta_t^{\hat{d}_1-d_0}\varepsilon_t\right)^2, \tag{2.47}$$

and, as a result,

$$T^{2d_0-1}W = T^{2d_0-1}t_{\vartheta_2}^2 \to \left(\frac{\mu_1-\mu_0}{\sigma}\right)^2 \frac{1-\lambda^{1-2d_0})}{(1-2d_0)\Gamma^2(1-d_0)}.$$

Note that the result in (2.17) does not change if μ_0 is estimated since, from the proof of Theorem 28, we have that

$$\vartheta_2 = \frac{T^{2d_0-1}\sum\limits_{t=[\lambda T]+1}^{T}\Delta_t^{d_0}1\Delta_t^{d_0}(y_t-\mu_0)}{T^{2d_0-1}\sum\limits_{t=[\lambda T]+1}^{T}(\Delta_t^{d_0}1)^2} - (\hat{\mu}_0 - \mu_0),$$

where the first term in the right hand side behaves as in (2.17) while the second term converges in probability to *zero.*

Part c)

Next, consider case (bl). First, we have that

$$T^{2d_0-1}F = \frac{T^{2d_0-1}(SSR_0 - SSR_1(\lambda))}{T^{-1}SSR_1(\lambda)} = T^{2d_0-1}D^R(1,2) + o_p(1),$$

where

$$D^R(1,2) = I^R + II^R,$$

with

$$\begin{aligned} I^R &= \left[T^{1/2}(\hat{d}_{0,1} - d_0)\right]^2 T^{-1} \sum_{1,i} F_{d,t}^2\left(\bar{\theta}_{0,i,t}\right) + \left[T^{1/2}(\hat{d}_{0,1} - d_0)\right]^2 T^{-1} \sum_{1,i} F_{d,t}^2\left(\bar{\theta}_{0,i,t}\right) \\ &+ \left[T^{1/2-d}\left(\hat{\mu}_{0,1} - \mu_0\right)\right]^2 T^{-1+2d} \sum_{1} F_{\mu,t}^2\left(\bar{\theta}_{0,i,t}\right) + \left[T^{1/2-d}\left(\hat{\mu}_{0,1} - \mu_1\right)\right]^2 T^{-1+2d} \sum_{2} F_{\mu,t}^2\left(\bar{\theta}_{0,i,t}\right) \end{aligned}$$

and

$$II^R = 2I^R$$

Next, from (2.37), note that

$$\begin{array}{c} \tilde{\mu} - \mu_0 \xrightarrow{p} (1 - \lambda^{1-2d})(\mu_0 - \mu_1) \\ \tilde{\mu} - \mu_1 \xrightarrow{p} \lambda^{1-2d}(\mu_0 - \mu_1) \end{array} .$$

Thus,

$$\begin{aligned} T^{1-2d_1} I^R \quad & \xrightarrow{p} \left(1 - \lambda^{1-2d_0}\right)^2 \left(\frac{\mu_1 - \mu_0}{\sigma}\right)^2 \frac{\lambda^{1-2d_0}}{(1-2d_0)\,\Gamma^2(1-d_0)} \\ & + \left(\lambda^{1-2d_0}\right)^2 \left(\frac{\mu_1 - \mu_0}{\sigma}\right)^2 \frac{1 - \lambda^{1-2d_0}}{(1-2d_0)\,\Gamma^2(1-d_0)} \\ = & \left(\frac{\mu_1 - \mu_0}{\sigma}\right)^2 \frac{\lambda^{1-2d_0}(1 - \lambda^{1-2d_0})}{(1-2d_0)\,\Gamma^2(1-d_0)}. \end{aligned}$$

Hence,

$$T^{2d-1}F = T^{2d_0-1} D^R(1,2)^2 \xrightarrow{p} \left(\frac{\mu_1 - \mu_0}{\sigma}\right)^2 \frac{\lambda^{1-2d_0}(1 - \lambda^{1-2d_0})}{(1-2d_0)\,\Gamma^2(1-d_0)}.$$

Chapter 3

Linear Trends, Fractional Trends and Initial Conditions

Abstract. This paper analyzes efficient estimation of linear trends in the long memory context. First, we consider the case of a long memory Type II error process and show that a generalized least squares (GLS) estimator that corrects the serial correlation of the error term is efficient. Second, we take into account an initial condition which bridges the two alternative definitions of long memory, Type I and Type II. In this case, a weighted least estimator (WLS), which is the efficient estimator for Type I, outperforms the GLS even for short initial conditions. It reaches efficiency when the initial condition becomes more remote. Consequently, the choice between the two estimators depends on the presence and the length of an initial condition. In order to illustrate the methodology, we estimate the GDP growth rates of three countries and test whether these rates are positive.

3.1 Introduction

Many economic time series have a trending behavior that can be described by models including a linear time trend. Estimating the slope of the trend provides information on the average growth of the series. For testing the hypothesis of positive growth we need asymptotically valid inference. This inference is more precise when we apply an efficient estimator. In order to determine which estimator is efficient, we have to take into account the persistence in the stochastic component described by additive long memory noise, nesting as special case the unit root situation, as well as the assumption about the pre-sample history.

In the Long-Memory literature, there exist two different characterizations of $I(d)$ processes: Type I and Type II. For the Type I specification, the estimation of deterministic trends is already thoroughly analyzed (Yajima (1988, 1991), Deo and Hurvich (1998), Hosking (1996)). Dahlhaus (1995) proposes an efficient weighted least squares estimator (WLS) for this case. However, the efficiency gains are very small. For the Type II specification, the efficient estimation of linear trends has not been explored yet. In this paper, we complement this literature by analyzing the efficient estimation for the Type II case. We show that in this case efficiency gains from using a properly defined generalized least squares (GLS) are large. Besides, we introduce an initial condition, which bridges the Type II and the Type I specification of long memory.

Canjels and Watson (1997) analyze trend estimation in the autoregressive unit root environment. They consider local-to-unity asymptotics with $\rho = 1 - c/T$, where T is the sample size, and show that asymptotic distributions and the efficiency depends on the parameter c, the presence and the length of an initial condition. In particular, efficiency increases in c and in the length of the initial condition. We extend the literature of efficient estimation of trends in the presence of initial conditions from the standard unit root context into the (nonstationary) long-memory context. We define an initial condition as $[\kappa T]$ pre-sample observations, where $\kappa \geq 0$. The parameter κ is the relative length of the initial condition to the observed sample. We analyze the effect of κ on the efficiency of the estimators by comparing GLS, WLS, first difference (FD) and ordinary least squares (OLS) to an efficient but infeasible maximum likelihood estimator (MLE).

We consider the memory parameter d lying in different intervals, $1/2 < d < 3/2$, $d = 1/2$ and $0 < d < 1/2$. The first corresponds to nonstationary LM and nests the unit root case for $d = 1$. We analyze the effect of the pre-sample history on the efficiency of the estimators. We show that for any memory $d \in (0, 3/2)$ the GLS estimate is efficient with large efficiency gains compared to OLS, especially for $d > 1$, in contrast to the Type I memory context, where potential efficiency gains are rather small. However, in the presence of an initial condition ($\kappa > 0$), the GLSE is not efficient anymore. In this case, we show that Dahlhaus' (1995) WLSE gets very close to the efficient MLE even for short initial conditions and thus efficiency gains are very small. Hence, it is only important to know whether $\kappa = 0$ or $\kappa > 0$. The feasible GLSE and WLSE have the same asymptotic distribution as the respective infeasible ones since they only depend on the knowledge of d, which can be consistently estimated. The efficient MLE, on the other hand, requires the knowledge of κ and is hence infeasible. Therefore, in practice, choosing between the estimators depends on the assumption about presence and length of the initial condition. In general, the potential efficiency gains from using our GLSE are larger than the potential losses, especially for $d > 1$.

Finally, we apply the methodology to the GDP growth rates of three countries – USA, France and Canada – and illustrate the effect on inference of the assumption about the presence of a pre-sample history. This assumption about the presence of a pre-sample history might make it harder to reject some hypothesis since the confidence interval becomes wider. In particular, we find that the zero growth hypothesis of France cannot be rejected anymore.

In Section 3.2, we describe the setup and introduce the different estimators. In Section 3.3, we derive and compare the asymptotic distribution of the estimators for the case of a Type II process without initial condition ($\kappa = 0$), with initial condition ($\kappa > 0$), and for the case of a Type I process. In Section 3.4, we discuss the feasibility of the estimators. In Section 3.5, we apply the methodology to GDP growth rates of three countries. Finally, in Section 3.6, we conclude. Lemmata and additional Propositions, which are needed for the analysis, are provided in Appendix A. The proofs are collected in Appendix B.

3.2 Preliminaries

In the Long-Memory literature, there exist two different characterizations of $I(d)$ processes: Type I and Type II. The Type I is differently defined depending on whether $d < 1/2$ or $d > 1/2$. In the former case, we define the process as an infinite moving average of short-memory innovations; in the latter, we define the process as a partial sum of Type I $I(d-1)$ terms, in a recursive way until memory is less than $1/2$. In the Type II specification, the process is defined as a $d-$fractionally integrated sum of $I(0)$ terms, truncated for $t \leq 0$, for any d (Marinucci and Robinson, 1999).

In this paper, we consider error processes following (i) Type II specification without initial condition ($\kappa = 0$), (ii) Type II specification with initial condition ($\kappa > 0$), and (iii) Type I specification, respectively. For (i), we assume that the series can be represented as

$$\begin{aligned} y_t &= \mu + \beta t + u_t, \text{ with} \\ u_t &= \Delta_{t+[\kappa T]}^{-d} v_t = \Delta_t^{-d} v_t + \xi_t I\left(\kappa > 0\right), \end{aligned} \tag{3.1}$$

where $\mu + \beta t$ is the deterministic component and v_t is a I(0) process. The error term u_t consists of the component $\Delta_t^{-d} v_t$ truncated to the sample and the initial condition $\xi_t = \Delta_{t+[\kappa T]}^{-d} v_t = \sum_{j=t}^{[\kappa T]+t-1} \pi_j(-d) v_{t-j}$, reflecting the sum of $[\kappa T]$ pre-sample terms. Δ_s^{-d} denotes the truncated fractional filter and

$$\left\{ \pi_j\left(-d\right) = \frac{\Gamma\left(j+d\right)}{\Gamma\left(d\right)\Gamma\left(j+1\right)} \right\}_{j=1}^{s-1}$$

denotes the associated sequence of coefficients of the expansion of Δ_s^{-d}. This initial condition bridges Type I and Type II definitions and corresponds to a distant past initialization proposed by Shimotsu and Phillips (2006). In contrast to the standard unit root literature, where $\rho < 1$ implies stationarity, in our case, the initial condition enters the asymptotic distribution for any d. The impact of the initial condition is increasing with d (since the process becomes more persistent) and is decreasing (increasing) in t for $d < 1$ ($d > 1$). The parameter κ determines the speed with which the number of terms in the initial condition increases with T, measuring the extent of the pre-sample history on the current data. For $d > 1/2$, the initial condition ξ_t is of order $O_p(T^{d-1/2})$ for every t. The error term v_t is a linear process of *iid.*$(0,\sigma^2)$ errors ε_t,

$$\begin{aligned} v_t &= w(L)\varepsilon_t, \text{ with} \\ w(L) &= \sum_{i=0}^{\infty} w_i L^i \text{ and } \sum_{i=0}^{\infty} i|w_i| < \infty, \end{aligned}$$

and $f_v\left(0\right) = \frac{1}{2\pi} w^2\left(1\right) > 0$, with $w\left(0\right) = 1$. Without loss of generality, we assume $\sigma^2 = 1$.

For (ii), the initial condition $\xi_t = 0$, so that $u_1 = v_1$ implying that there is no pre-sample history. The case of a finite number of pre-sample terms,

$$\xi_t = \sum_{j=t}^{N+t-1} \pi_j\left(-d\right) v_{t-j},$$

where N is finite, provides similar asymptotic results as the case with $\kappa = 0$ (see Chung and Baillie, 1993).

For (iii), we use the errors v_t as in (i) and construct a Type I process, which is defined differently in the stationary ($d < 1/2$) and nonstationary ($1/2 < d < 3/2$) case. For $d < 1/2$,

$$u_t = \Delta_{\infty}^{-d} v_t = \sum_{j=0}^{\infty} \pi_j (-d) v_{t-j}. \tag{3.2}$$

We see clearly that our Type II process with initial condition lies between a Type I and a Type II process. In this case, for $\kappa \to \infty$, the process corresponds to a Type I process. For $1/2 < d < 3/2$, we define the process as a partial sum of stationary $I(d-1)$ processes:

$$u_t = \sum_{k=1}^{t} \Delta_{\infty}^{-d+1} v_k = \sum_{k=1}^{t} \sum_{j=0}^{\infty} \pi_j (1-d) v_{k-j}. \tag{3.3}$$

Remark 1. *In the unit root literature, there is a different definition of initial conditions, $y_0 = O_p(\sqrt{T})$ or $y_0 = O(\sqrt{T})$ (see Elliott (1999) and Müller and Elliott (2003) respectively). While in the AR(1) unit root context the former definition is similar to the definition $y_0 = \sum_{j=1}^{[\kappa T]} \pi_j (-1) \varepsilon_{1-j}$, in our context, the two definitions are different. We could consider*

$$\begin{aligned} \Delta_{t+1}^{d} y_t &= \varepsilon_t \\ y_t &= \left(1 - \Delta_{t+1}^{d}\right) y_t + \varepsilon_t = \sum_{j=1}^{t-1} \pi_j (d) y_{t-j} + \pi_t y_0 + \varepsilon_t, \end{aligned}$$

where $y_0 = O_p\left(T^{d-1/2}\right)$ or $y_0 = O\left(T^{d-1/2}\right)$. *This alternative definition appears to be less natural in our context but it still could be analyzed along similar lines.*

In this paper, we analyze different estimators for the linear trend term β in (3.1): ordinary least squares (OLS), first-difference estimation (FD), generalized least squares (GLS), Dahlhaus' (1995) weighted least squares (WLS) and the maximum likelihood estimator (MLE).

For the OLS estimator of equation (3.1), y_t is regressed on a constant term and t. This leads to

$$\hat{\beta}_{OLS} = \frac{\frac{1}{T} \sum_{t=1}^{T} t y_t - \left(\frac{1}{T} \sum_{t=1}^{T} t\right) \left(\frac{1}{T} \sum_{t=1}^{T} y_t\right)}{\frac{1}{T} \sum_{t=1}^{T} t^2 - \left(\frac{1}{T} \sum_{t=1}^{T} t\right)^2}.$$

The FD estimator results from applying a FD operator on equation (3.1)

$$\Delta y_t = \beta + \Delta u_t, t \geq 2, \tag{3.4}$$

and regressing Δy_t on 1. Thus,

$$\hat{\beta}_{FD} = \frac{1}{T-1} \sum_{t=2}^{T} \Delta y_t.$$

Next, both GLS and WLS depend on the knowledge of the memory parameter d. We assume d to be known, and concentrate on the estimation of the trend coefficient. The case of unknown d, will be discussed in Section 4. For the GLS estimator, we apply the truncated filter Δ_t^d on equation (3.1),

$$\Delta_t^d y_t = \mu \Delta_t^d 1 + \beta \Delta_t^d t + v_t + \Delta_t^d \xi_t I\left(\kappa > 0\right), \tag{3.5}$$

where $\Delta_t^d \xi_t I\left(\kappa > 0\right) = \Delta_t^d \sum_{j=t}^{[\kappa T]+t-1} \pi_j(-d) v_{t-j} I\left(\kappa > 0\right)$. For $\kappa = 0$, the resulting residuals v_t in (3.5) are short memory. For $\kappa > 0$, in contrast to the GLS in the standard unit root context, the variances of the initial condition differ for different t due to the truncation. The GLS estimator can be written as

$$\hat{\beta}_{GLS} = \frac{(q_{11}^T s_2^T - q_{12}^T s_1^T)}{(q_{11}^T q_{22}^T - (q_{12}^T)^2)}, \tag{3.6}$$

where the deterministic terms

$$q_{11}^T = \sum_{t=1}^{T} (\Delta_t^d 1)^2, \; q_{12}^T = \sum_{t=1}^{T} (\Delta_t^d 1)(\Delta_t^d t) \text{ and } q_{22}^T = \sum_{t=1}^{T} (\Delta_t^d t)^2 \tag{3.7}$$

and the stochastic terms

$$s_1^T = \sum_{t=1}^{T} (\Delta_t^d 1) \Delta_t^d y_t \text{ and } s_2^T = \sum_{t=1}^{T} (\Delta_t^d t) \Delta_t^d y_t.$$

This GLS estimator corresponds to the conditional sum-of-squares (CSS) estimator[1]. For simplicity, this estimator corrects only the serial correlation associated with the fractional integration, while it ignores the $I(0)$ serial correlation associated with $w(L)$. From Grenander-Rosenblatt (1957)[2], OLS and GLS treatments of $w(L)$ are asymptotically equally efficient in our case. Hence, ignoring the $I(0)$ serial cor-

[1] Chung and Baillie (1993) analyze this estimator for a Type I process with fixed initial observations and $d < 1/2$ for the estimation of the mean, the parameter d and the ARMA coefficients. Ling and Li (2001) extend it to nonstationary Type II error processes. Gil-Alana (2008) uses this estimator for a trending Type II process with breaks in determinsitics and memory. Beran (1995), Robinson (1996) and Hualde and Robinson (2011) are further references for the CSS estimation of the memory parameter.

[2] See also Chapter 10 in Palma (2007).

relation has no asymptotic effect on the estimators of the linear trend. Note that all filters are truncated, so that $\Delta_t^d 1$ does not vanish[3]. The GLS estimator is then the OLS estimator applied to equation (3.5), regressing $\Delta_t^d y_t$ on $\Delta_t^d 1$ and $\Delta_t^d t$.

Dahlhaus (1995) proposes a WLS estimator that is efficient for polynomial trend regression when the error process follows a Type I process with a memory $d < 1/2$. This estimator is directly extendable to the nonstationary case ($1/2 < d < 3/2$). The WLS estimators for equations (3.1) and (3.4) coincide. For simplicity, we employ the latter. From Dahlhaus (1995), we obtain for this case the regressor $X_i = \varphi_1(x) = 1$ and the corresponding weighting matrix

$$w_d(x) = (x)^{1-d}(1-x)^{1-d}.$$

This leads to Dahlhaus' WLS,

$$\hat{\beta}_{WLS} = \frac{\sum_{t=2}^{T} w_d\left(\frac{t}{T+1}\right)\Delta y_t}{\sum_{t=2}^{T} w_d\left(\frac{t}{T+1}\right)}. \tag{3.8}$$

Finally, we construct the Gaussian MLE. Sowell (1992) analyzes a Full MLE for a Type I error process with $d < 1/2$ for the estimation of d and the ARMA coefficients but without deterministic terms. For a Type II process with initial condition, the MLE is asymptotically efficient. However, it is computationally very demanding and depends on the knowledge of both d and κ. This estimator minimizes the sum of the squared difference between y_t and its best linear one-step predictor (BLP), standardized by its forecast error variance. Since the process is nonstationary, we apply the innovations algorithm (Brockwell and Davis (1991), Prop. 5.2.2). In particular, the (nonstationary) series

$$y_t = \mu + \beta t + u_t$$

is predicted by:

$$\hat{y}_t = \hat{\mu} + \hat{\beta}t + \hat{u}_t$$

where $\hat{u}_t$ is obtained in the innovations algorithm. The BLP of u_t is

$$\hat{u}_t = \sum_{i=1}^{t-1}\theta_{t-1,i}(u_{t-i} - \hat{u}_{t-i}) = \sum_{i=1}^{t-1}\theta_{t-1,i}\left(y_{t-i} - \hat{\mu} - \hat{\beta}(t-i) - \hat{u}_{t-i}\right).$$

[3] See Gil-Alana (2008) for graphical illustrations of $\Delta_t^d 1$ and $\Delta_t^d t$ for different values of d.

For the innovations algorithm, we need $\{u_t\}$ to have zero mean and the covariance matrix $[\sigma_u(i,j)]_{i,j=1}^T$, with elements $E(u_i u_j) = \sigma_u(i,j)$, to be non-singular for any T. Asymptotically, the former results from subtracting the estimated deterministic component from y_t. For the latter, the nonsingular covariance matrix is

$$\sigma_u(i,j) = \sum_{k=0}^{[\kappa T]+i} \pi_k(-d)\, \pi_{k+|j-i|}(-d)$$

for $\kappa \geq 0$. This MLE takes into account the pre-sample history but, equally as the GLS, neglects the $I(0)$ serial correlation, again relying on the Grenander and Rosenblatt (1957) result. We could also include this correlation modifying the covariance matrix $[\sigma_u(i,j)]_{i,j=1}^T$. Finally, the one-step predictor is obtained by the innovations algorithm (see Lemma 15 in Appendix A). Consequently, given d and κ, the MLE is

$$\arg\min_{\mu,\beta} Q(\mu,\beta,d,\kappa) = \arg\min_{\mu,\beta} \sum_{t=1}^{T} \left(\frac{e_t(\mu,\beta,d,\kappa)}{\sqrt{v_t^*(d,\kappa)}} \right)^2 \tag{3.9}$$

where $e_t(\mu,\beta,d,\kappa) = y_t - \mu - \beta t - \hat{u}_t$ and $v_t^*(d,\kappa)$ is the forecast error variance $E(u_t - \hat{u}_t)^2$ obtained in the algorithm. In the argument of (3.9), we substitute recursively $\{\hat{u}_j\}_{j=1}^t$ and we obtain the difference of a linear function of y_t, $y_t^* = y_t - \sum_{k=1}^{t-1} \phi_k y_{t-k}$ (nonlinear in θ), and functions of the constant term and t, ψ_{1t} and ψ_{2t} respectively:

$$x_t^* = y_t^* - \psi_{1,t}\mu - \psi_{2,t}\beta.$$

Lemma 42 in Appendix A indicates how to obtain each of the terms from the terms of the innovations algorithm. Given the data series y_t we construct y_t^*, ψ_{1t} and ψ_{2t} for $t = 1, ..., T$ and regress $y_t^*/\sqrt{v_t^*}$ on $\psi_{1t}/\sqrt{v_t^*}$ and $\psi_{2t}/\sqrt{v_t^*}$. Therefore, the MLE estimator can be written as (3.6), where now

$$q_{11}^T = \sum_{t=1}^{T} \psi_{1t}^2/v_t^*,\ \ q_{12}^T = \sum_{t=1}^{T} \psi_{1t}\psi_{2t}/v_t^* \text{ and } q_{22}^T = \sum_{t=1}^{T} \psi_{2t}^2/v_t^* \tag{3.10}$$

and

$$s_1^T = \sum_{t=1}^{T} \psi_{1t} y_t^*/v_t^* \text{ and } s_2^T = \sum_{t=1}^{T} \psi_{2t} y_t^*/v_t^*.$$

3.3 Asymptotic Distributions of the Estimators

In this section we derive and compare the asymptotic distribution of the estimators. We analyze the impact of the length of the initial condition (κ) on the estimators. We consider the cases of nonstationary, $1/2 < d < 3/2$, $d = 1/2$, and (asymptotically) stationary $0 < d < 1/2$, separately.

First, we consider a memory $1/2 < d < 3/2$. Let

$$V_1(d,\kappa) = \frac{\int_0^\kappa (s+s^2)^{d-1} ds}{\Gamma^2(d)}$$

and

$$V_2(d,\kappa) = (2d-1)\int_0^1 \int_0^t \left(s-\frac{1}{2}\right)\left(t-\frac{1}{2}\right)\int_0^{\kappa+s} ((\kappa+t)-x)^{d-1}((\kappa+s)-x)^{d-1} dx ds dt. \tag{3.11}$$

Further, let

$$\begin{aligned}
V_3(d,\kappa) &= \frac{(3-2d)^2\Gamma^4(2-d)\Gamma^4(d)}{\Gamma^2(2d-1)} A(d,\kappa)\text{, where} && (3.12)\\
A(d,\kappa) &= \lim_{T\to\infty} T^{2d-3}\sum_{j=0}^{[\kappa T]}\left(\sum_{t=1}^{T}\chi_t^T(d)\sum_{k=0}^{t-1}\pi_k(d)\,\pi_{t-k+j}(-d)\right)^2 \text{ and} && (3.13)\\
\chi_t^T(d) &\equiv q_{11}^T\pi_{t-1}(d-2) - q_{12}^T\pi_{t-1}(d-1), && (3.14)
\end{aligned}$$

where q_{11}^T and q_{12}^T are defined in (3.7). Next, let

$$\begin{aligned}
V_{41}(d) &= \lim_{T\to\infty} T^{1-2d}\sum_{t=1}^{T}\left(\sum_{k=0}^{T-t}\pi_k(1-d)\,w_d\left(\frac{t+k}{T+1}\right)\right)^2 \text{ and} && (3.15)\\
V_{42}(d,\kappa) &= \frac{1}{\Gamma^2(d-1)}\int_0^\kappa\left(\int_0^1 s^{1-d}(1-s)^{1-d}(s+k)^{d-2} ds\right)^2 dk. && (3.16)
\end{aligned}$$

Finally, let

$$V_5(d,\kappa) = \frac{\lim_{T\to\infty} T^{d-3/2}\sum_{t=1}^{T}\left(q_{11}^T\psi_{2t} - q_{12}^T\psi_{1t}\right)^2}{\left(\lim_{T\to\infty} T^{2d-3}\left(q_{11}^T q_{22}^T - (q_{12}^T)^2\right)\right)^2},$$

where q_{11}^T, q_{12}^T, ψ_{1t} and ψ_{2t} are defined in and below (3.10). Then, Theorem 34 discusses the asymptotic distributions of the estimators in function of the parameters d and κ. Let β_0 denote the true trend parameter in (3.1).

Theorem 34 *(Asymptotic distribution for a Type II process with initial condition*

κ and with $1/2 < d < 3/2$.)

(a) $T^{3/2-d}(\hat{\beta}_{FD}-\beta_0) \xrightarrow{d} N\left(0, w(1)^2\left[\frac{(\kappa+1)^{2d-1}+\kappa^{2d-1}}{\Gamma^2(d)(2d-1)} - 2V_1(d,\kappa)\right]\right)$.

(b) $T^{3/2-d}(\hat{\beta}_{OLS}-\beta_0) \xrightarrow{d} N\left(0, w(1)^2\frac{144V_2(d,\kappa)}{\Gamma^2(d)(2d-1)}\right)$.

(c) $T^{3/2-d}(\hat{\beta}_{GLS}-\beta_0) \xrightarrow{d} N\left(0, w(1)^2\left[(3-2d)\Gamma^2(2-d)+V_3(d,\kappa)\right]\right)$.

(d) $T^{3/2-d}(\hat{\beta}_{WLS}-\beta_0) \xrightarrow{d} N\left(0, w(1)^2\frac{\Gamma^2(4-2d)}{\Gamma^4(2-d)}\left[V_{41}(d)+V_{42}(d,\kappa)\right]\right)$.

(e) $T^{3/2-d}(\hat{\beta}_{MLE}-\beta_0) \xrightarrow{d} N\left(0, w(1)^2 V_5(d,\kappa)\right)$.

Remark 2. *For $\kappa = 0$, the expressions of the asymptotic variances simplify considerably. The terms $V_1(d,\kappa)$, $V_3(d,\kappa)$ and $V_{42}(d,\kappa)$ vanish and $V_2(d,\kappa)$ and $V_5(d,\kappa)$ simplify, e.g. $V_5(d,\kappa) = (3-2d)\Gamma^2(2-d)$.*

Remark 3. *For $\kappa > 0$, we do not find an exact asymptotic expression for these terms and have to rely on numerical evaluations. We are not able to obtain a closed expression for the asymptotic variance of the MLE because of its recursive nature. In the following, we calculate the terms $V_1(d,\kappa)$, $A(d,\kappa)$, $V_{41}(d)$ for sample sizes of $T = 10,000$.*

Remark 4. *As a special case – for $d = 1$ – FD corresponds to GLS leading to a standard result (Phillips and Durlauf, 1988). In this case, the initial condition does not have an effect on the asymptotic distribution of any of the estimators, leading to Corollary 1, which follows from substituting $d = 1$ in Theorem 1.*

Corollary 35 *(Asymptotic distribution for errors with a unit root ($d = 1$))*
(a) For FD, GLS/MLE, WLS:

$$T^{1/2}(\hat{\beta}_{\vartheta}-\beta_0) \xrightarrow{d} N(0, w(1)^2),\ \vartheta = \{FD, GLS, MLE, WLS\}.$$

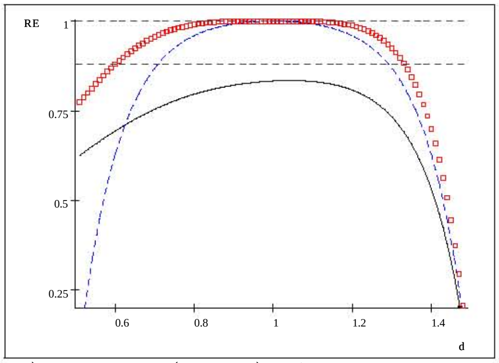
(b) For OLS:

$$T^{1/2}(\hat{\beta}_{OLS}-\beta_0) \xrightarrow{d} N\left(0, w(1)^2 6/5\right).$$

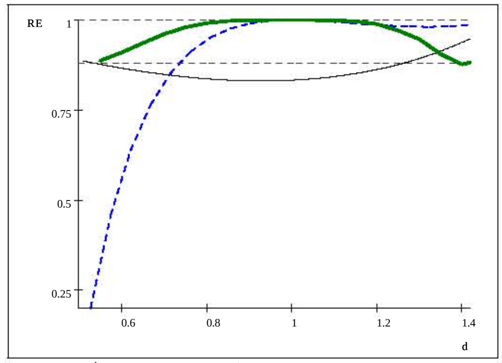
For completeness, Theorem 36 provides the asymptotic distributions of the estimators for the specific Type I process (3.3). These distributions are derived in Hosking (1996), Deo and Hurvich (1998) in form of a numerical integral, Dahlhaus (1995). The GLS and MLE ones are derived in the Appendix. Let

$$V_6(d) = \lim_{T\to\infty} T^{2d-3}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{k=0}^{T-t}\chi_{t-1}^{T}(d)\pi_k(d)\right)\left(\sum_{k=0}^{T-s}\chi_{t-1}^{T}(d)\pi_k(d)\right)\gamma'_{s,t},$$

Figure 3-1: **: Asymptotic Relative Efficiency (ARE) for 1/2<d<3/2**



a) Type II ($\kappa = 0$): ARE to GLS



b) Type I: ARE to WLS

GLS (green thick line), WLS (red boxes), FD (blue dashed line), OLS (black thin line)

where $\chi_t^T(d)$ is defined in (3.14) and

$$\gamma'_{s,t} = \frac{\Gamma(3-2d)}{(2d-1)\Gamma(d)\Gamma(2-d)}\frac{1}{2}\left(s^{2d-1} + t^{2d-1} - |t-s|^{2d-1}\right).$$

Theorem 36 *(Asymptotic distribution for a Type I with $1/2 < d < 3/2$)*

(a) $T^{3/2-d}(\hat{\beta}_{FD} - \beta_0) \xrightarrow{d} N\left(0, w(1)^2 \frac{1}{(2d-1)}\frac{\Gamma(3-2d)}{\Gamma(2-d)\Gamma(d)}\right).$

(b) $T^{3/2-d}(\hat{\beta}_{OLS} - \beta_0) \xrightarrow{d} N\left(0, w(1)^2 \frac{18}{d(2d+3)(2d+1)}\frac{\Gamma(3-2d)}{\Gamma(2-d)\Gamma(d)}\right).$

(c) $T^{3/2-d}(\hat{\beta}_{GLS} - \beta_0) \xrightarrow{d} N(0, w(1)^2 V_6(d)).$

(d) $T^{3/2-d}(\hat{\beta}_{WLS} - \beta_0) \xrightarrow{d} N\left(0, w(1)^2 \frac{\Gamma^2(3-2d)(3-2d)}{\Gamma^2(2-d)}\right).$

(e) $T^{3/2-d}(\hat{\beta}_{MLE} - \beta_0) \xrightarrow{d} N\left(0, w(1)^2 \frac{\Gamma^2(3-2d)(3-2d)}{\Gamma^2(2-d)}\right).$

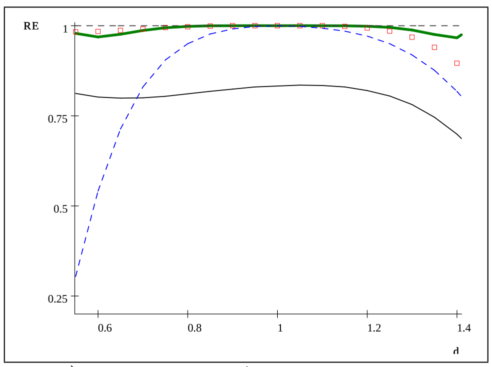
Next, we compare the asymptotic variances of the different estimators for the three cases: Type II ($\kappa = 0$), Type I, and Type II with an initial condition with $\kappa > 0$. For $\kappa = 0$, Figure 3-1a) shows the relative efficiency of the FD, OLS and WLS to GLS/MLE for different values of $d \in (1/2, 3/2)$. OLS is more efficient than FD for values of d smaller than 0.63; for values above 0.63 the FD is more efficient. The relative efficiency of FD to OLS is lower than the one under Type I (Figure 3-1b)) for any value except $d = 1$. The WLS is more efficient than the OLS and the FD and its efficiency increases with d departing from *one*. As expected, the GLS/MLE is the efficient estimator and is more efficient than its alternatives for any $d \neq 1$. All three alternative estimators become highly inefficient relative to the GLS if $d \to 3/2$. This stands in stark contrast to the gains from efficient estimation (Dahlhaus' WLS, 1995) in Type I. Figure 3-1b) shows that the gains from efficient estimation are rather small. The relative efficiency for $d > 1$ exceeds 98% as found in Yajima (1988).

Comparing the estimators under Type II and Type I, we find two big differences (especially for $d > 1$): first the variances of the estimators are much higher in the Type I definition and, second, the gains from efficient estimation are much higher in the Type II definition. Shimotsu and Phillips (2006) discuss that the essential difference between both definitions is one of initialization. They show that the process under the Type I specification can be written as $u_t = u_0 + \xi_{t0}(d) + \Delta_t^{-d} v_t$ where the additional term $\xi_{t0}(d)$ has the same stochastic order as the third term. Therefore, it adds uncertainty to the process. None of the estimators can control this noise causing their bad performance, especially, when the term is of higher order ($d > 1$). In contrast, under the Type II definition, while all estimators perform much better due to the smaller noise, the GLS adapts perfectly to the truncated fractionally integrated structure of the process. This implies its clearly superior performance.

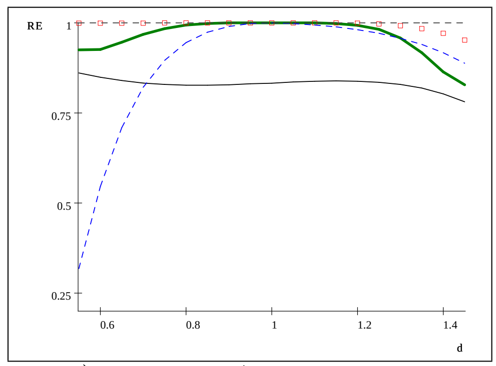
For the Type II with initial condition, Figure 3-2 shows the relative efficiency of FD, OLS, WLS, and GLS to the MLE as a function of d for $\kappa = 0.01,\ 0.1,\ 1$ and 5. Interestingly, especially for $d > 1$, the GLS is affected by the initial condition. This contrasts to the case without initial condition. The reason is that for the case without initial condition, $\Delta_t^d u_t = v_t$ and for the case with initial condition, $\Delta_t^d u_t = v_t + \Delta_t^d \sum_{j=t}^{[\kappa T]+t-1} \pi_j(-d) v_{t-j}$. This additional term increases the asymptotic variance. This efficiency loss implies a drawback of the GLS estimator since in practice, we do not know whether $\kappa = 0$ or whether $\kappa > 0$. For a very small $\kappa = 0.01$ and for $d > 1$, the GLS is still more efficient than the WLS. As κ increases, WLS gets close to the efficient MLE and outperforms the alternatives. GLS is more efficient than FD and OLS except for d close to $3/2$. Finally, FD and OLS slightly gain efficiency for an increasing length of the initial condition. This is not surprising since when κ continue increasing, the results get closer to the ones under Type I errors. Comparing Figure 3-1a) with Figure 3-2 shows that the introduction of an initial condition does not change the qualitative pattern between OLS and FD. Finally, even already for a very small $\kappa = 0.01$, the graph differs completely from the one of $\kappa = 0$. The graphs for $\kappa = 1$ and $\kappa = 5$ (Figure 3-2c) and d)) look virtually the same. Consequently, the critical length of the initial condition is rather short.

A memory of $d = 1/2$ constitutes an interesting special case, since at this point the process becomes nonstationary, while for $d < 1/2$ the process has long memory but is (asymptotically) stationary. For a Type I process, it is known that there is

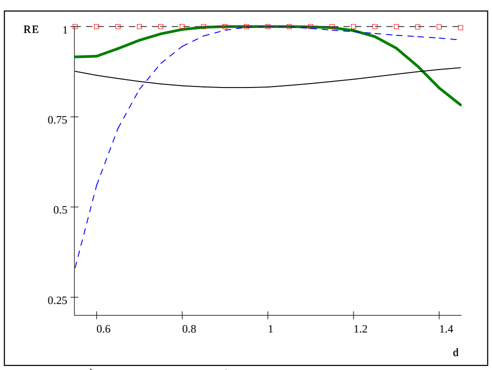
Figure 3-2: **: Asymptotic Relative Efficiency (ARE) for Type II with IC for 1/2<d<3/2**



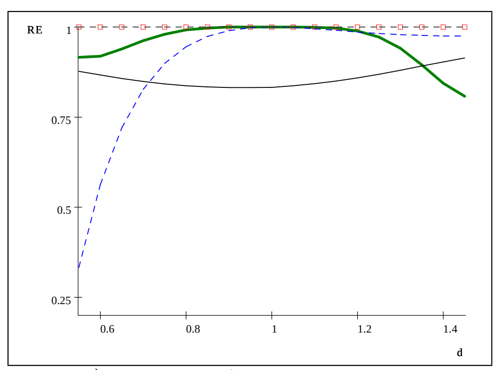
a) $\kappa = 0.01$: ARE to MLE



b) $\kappa = 0.1$: ARE to MLE



c) $\kappa = 1$: ARE to MLE



d) $\kappa = 5$: ARE to MLE

GLS (green thick line), WLS (red boxes), FD (blue dashed line), OLS (black thin line) a discontinuity affecting the rate of convergence (see Theorem 8 in Hosking, 1996). Proposition 37 illustrates that, for a Type II process, only the FD estimator has such a discontinuity.

Proposition 37 *(Asymptotic distribution for a Type II process with initial condition $\kappa > 0$ and with $d = 1/2$)*

(a) $\quad T\,(lnT)^{-1/2}\,(\hat{\beta}_{FD} - \beta_0) \xrightarrow{d} N\left(0, w(1)^2 \frac{1 + I\,(\kappa > 0)}{\pi}\right).$

(b) $\quad$ *The asymptotic distributions of OLS,GLS,WLS and MLE correspond to the ones in Theorem 1 with $d = 1/2$.*

Remark 5. *The rate of convergence of the FD estimator $T\,(lnT)^{-1/2}$ is lower than the one of its alternatives and lies in a discontinuous way between the one for $d < 1/2$ (rate T) and the one for $d > 1/2$ (rate $T^{3/2-d}$). For $d = 1/2$, the variance of the FD corresponds for a Type II process with an initial condition of any length κ to the one of a Type I process. For the Type I process, the FD estimator has a variance that is the double of the one of a Type II process (see Theorem 8 in Hosking, 1996). The difference reflects the fact that in contrast to the Type I the process is truncated for the Type II.*

Remark 6. *The introduction of an initial condition also creates an additional term depending on κ in the variances of the OLS, GLS and WLS estimators. Their variances lie for an initial condition of any length κ in an continuous way between the ones for $d < 1/2$ and $d > 1/2$. Thus, they correspond to the limit variances for $d \to 1/2$ from above (Theorem 1) and below (Theorem 4) and can be found in Figures 3-1a) and 3-2.*

For $0 \leq d < 1/2$, the two definitions of fractional processes are asymptotically equivalent in the sense that

$$\sum_{k=0}^{\infty} \pi_k\,(-d)\,u_{t-k} = \sum_{k=0}^{t-1} \pi_k\,(-d)\,u_{t-k} + O_p\left(t^{2d-1}\right),$$

where the second term converges to *zero* as $t \to \infty$ (Marinucci and Robinson, 1999). However, this does not imply that, for a Type II process, the variances of FD and OLS estimator correspond to the ones of Type I. In fact, the implied fractional Brownian Motions are of Type II and Type I respectively (see Davidson and Hashizade, 2009, for a comparison). Since for $d < 1/2$, the initial condition bridges the Type I and Type II processes, the variance for a process with initial condition will also lie between the ones of these two processes. In particular, Theorem 38 provides the asymptotic distributions under Type II. First, let

$$V_7(d,\kappa)=\frac{144}{\Gamma^2(d)}\int_0^{\kappa}\left(\int_0^1\left(s-\frac{1}{2}\right)(k+s)^{d-1}ds\right)^2 dk.$$

Further, let

$$\begin{aligned}
V_{81}(d) &= 4\Gamma^2(2-d)(3-2d)(d-1)^2 \text{ and} \\
V_{82}(d,\kappa) &= \left(\frac{\Gamma^4(2-d)\,4(1-2d)(3-2d)}{4(1-d)^2-(1-2d)(3-2d)}\right)^2 A(d,\kappa),
\end{aligned}$$

where $A(d,\kappa)$ is defined as (3.13) with a rate $T^{3d-5/2}$ instead of T^{2d-3}.

Theorem 38 *(Asymptotic distribution for a Type II process with initial condition $\kappa>0$ and $d<1/2$)*

(a) $T(\hat{\beta}_{FD}-\beta_0)$ *converges in distribution to a random variable with zero mean and variance* $w(1)^2\left((1-I(\kappa>0))+(1+I(\kappa>0))\frac{\Gamma(1-2d)}{\Gamma^2(1-d)}\right)$ *with the distribution depending on the distribution of the underlying errors* $\varepsilon's$.

(b) $T^{3/2-d}(\hat{\beta}_{OLS}-\beta_0)\xrightarrow{d}N\left(0,w(1)^2\left[\frac{36(2d^3-d^2+1)}{(2d+1)(2d+3)\Gamma^2(d+2)}+V_7(d,\kappa)\right]\right).$

(c) $T^{3/2-d}(\hat{\beta}_{GLS}-\beta_0)\xrightarrow{d}N\left(0,w(1)^2\left[V_{81}(d)+V_{82}(d,\kappa)\right]\right).$

(d) $T^{3/2-d}(\hat{\beta}_{WLS}-\beta_0)\xrightarrow{d}N\left(0,w(1)^2\frac{\Gamma^2(4-2d)}{\Gamma^4(2-d)}\left[V_{41}(d)+V_{42}(d,\kappa)\right]\right).$

(e) $T^{3/2-d}(\hat{\beta}_{MLE}-\beta_0)\xrightarrow{d}N\left(0,w(1)^2V_5(d,\kappa)\right).$

Remark 7. *The terms $V_{41}(d)$, $V_{42}(d,\kappa)$ and $V_5(d,\kappa)$ correspond to the ones in Theorem 1.*

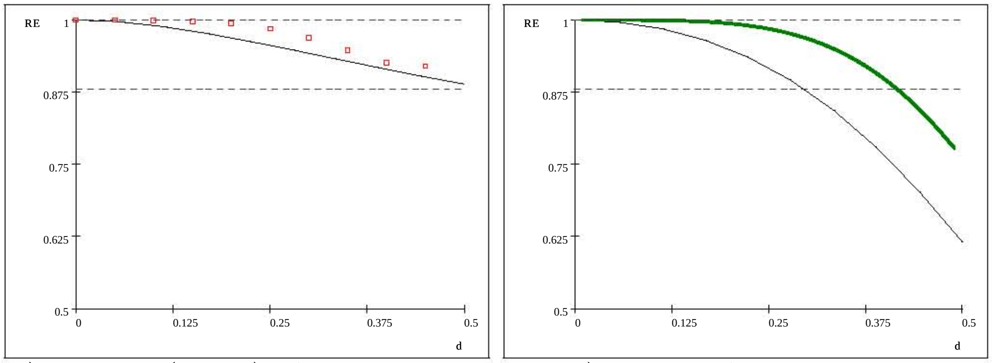
Remark 8. *The rate of convergence of FD is lower than the one of the alternative estimators and does not depend on d.*

Remark 9. *The GLS coincides with the MLE estimator and is efficient.*

These variances clearly differ from the ones of a Type I process in Theorem 39. Let

$$V_{10}(d)=\lim_{T\to\infty}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{k=0}^{T-t}\chi_{t-1}^{T}(d)\,\pi_k(d)\right)\left(\sum_{l=0}^{T-s}\chi_{t-1}^{T}(d)\,\pi_l(d)\right)\gamma_{s-t},$$

Figure 3-3: **: Asymptotic Relative Efficiency (ARE) for d<1/2**



a) Type II ($\kappa = 0$): ARE to GLS   b) Type I: ARE to WLS

GLS (green thick line), WLS (red boxes), OLS (black thin line)

where $\chi_{t-1}^{T}(d)$ is defined in (3.14) and where now

$$\gamma_{s-t} = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \frac{\Gamma(|s-t|+d)}{\Gamma(1+|s-t|-d)}.$$

Theorem 39 *(Asymptotic distribution for a Type I process with memory d<1/2)*

(a) $T(\hat{\beta}_{FD} - \beta_0)$ *converges in distribution to a random variable with zero mean and variance* $w(1)^2 \, 2\dfrac{\Gamma(1-2d)}{\Gamma^2(1-d)}$ *with the distribution depending on the distribution of the underlying errors* $\varepsilon' s$.

(b) $T^{3/2-d}(\hat{\beta}_{OLS} - \beta_0) \xrightarrow{d} N\left(0, w(1)^2 \dfrac{18}{(2d+3)(2d+1)} \dfrac{\Gamma(3-2d)}{\Gamma(1+d)\Gamma(2-d)}\right)$

(c) $T^{3/2-d}(\hat{\beta}_{GLS} - \beta_0) \xrightarrow{d} N\left(0, w(1)^2 V_{10}\right)$

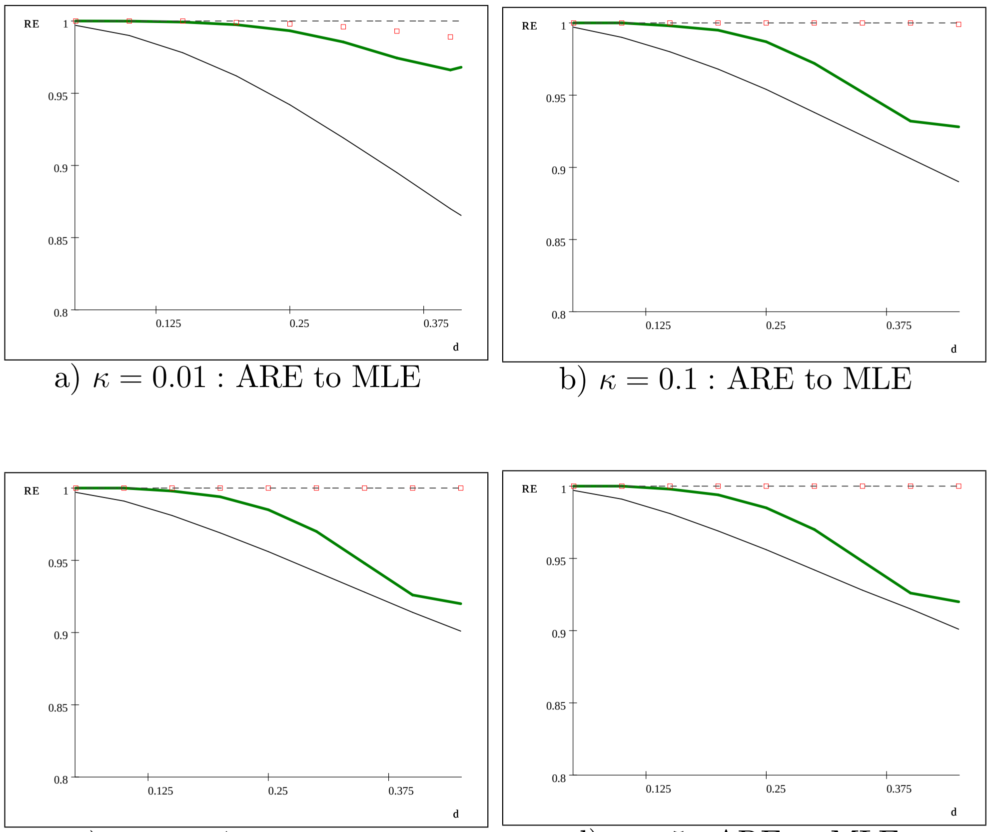
(d) $T^{3/2-d}(\hat{\beta}_{WLS} - \beta_0) \xrightarrow{d} N\left(0, w(1)^2 \dfrac{\Gamma^2(3-2d)(3-2d)}{\Gamma^2(2-d)}\right)$

(e) $T^{3/2-d}(\hat{\beta}_{MLE} - \beta_0) \xrightarrow{d} N\left(0, w(1)^2 \dfrac{\Gamma^2(3-2d)(3-2d)}{\Gamma^2(2-d)}\right)$

Figure 3-3b) displays the relative efficiency of the OLS and GLS to the efficient WLS for Type I. This confirms Yajima's (1988) result that for this memory region, even though OLS is inefficient, its efficiency lies above 88.8%. Consequently, for a Type I process, the Dahlhaus' (1995) WLS is not much more efficient than OLS.

For the Type II process, this efficiency bound does not hold and, consequently, the gains from efficient estimation are higher. Figure 3-3a) displays the relative efficiency of OLS and WLS to GLS/MLE for different values of d. The larger d is,

Figure 3-4: **: Asymptotic Relative Efficiency (ARE) for a Type II with IC for 0<d<1/2**



a) $\kappa = 0.01$: ARE to MLE

b) $\kappa = 0.1$: ARE to MLE

c) $\kappa = 1$: ARE to MLE

d) $\kappa = 5$: ARE to MLE

GLS (green thick line), WLS (red boxes), OLS (black thin line)

the less efficient relative to GLS become these two estimators, and the less efficient becomes OLS relative to WLS. For a Type II process, even between WLS and OLS, Yajima's efficiency bound does not hold.

Figure 3-4 shows the relative efficiency of the OLSE, WLSE and GLSE to the MLE as a function of $d < 1/2$ for a Type II with initial condition with $\kappa = 0.01,\ 0.1,\ 1$ and 5. The WLS is almost as efficient as the MLE already for $\kappa \geq 0.1$. Only for $\kappa = 0.01$ and d close to $1/2$, the MLE is more efficient. The relative efficiency of GLS to OLS decreases as κ increases, because the former gets less and the latter more efficient as κ increases.

In order to get some further indications about the performance of the different estimators, we determine the weights that each of them gives to each observation.

FD ($i = 1$), WLS ($i = 2$) and GLS ($i = 3$) are all linear estimators of Δy_t,

$$\frac{1}{T}\sum_{t=1}^{T}\omega_t^{(i)}\Delta y_t,$$

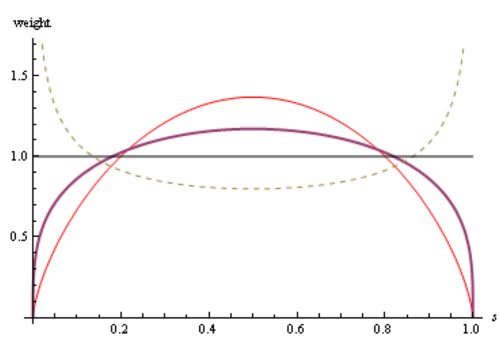
where

$$\begin{aligned}
\omega_t^{(1)} &= 1, \\
\omega_t^{(2)} &= \frac{\left(\frac{t}{T+1}\right)^{1-d}\left(1-\frac{t}{T+1}\right)^{1-d}}{\left(\frac{1}{T}\sum\limits_{t=1}^{T}\left(\frac{t}{T+1}\right)^{1-d}\left(1-\frac{t}{T+1}\right)^{1-d}\right)} \text{ and} \\
\omega_t^{(3)} &= \frac{\sum\limits_{i=0}^{T-t}\pi_i\left(d-1\right)\pi_{t+i-1}\left(d-2\right)}{\sum\limits_{i=2}^{T}\pi_i^2\left(d-1\right)\sum\limits_{i=2}^{T}\pi_i^2\left(d-2\right)},
\end{aligned}$$

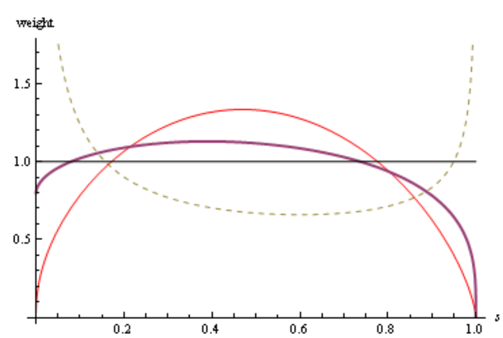
where we drop in $\omega_t^{(3)}$ some asymptotically negligible terms. For $\kappa = 0$, the MLE corresponds to the GLSE and for $\kappa > 0$, its weights depend on κ.

Figure 3-5 shows the asymptotic weights of the WLSE and GLSE for d=0.3 (thin line), 0.7 (thick line), and 1.3 (dashed line). For $d = 1$, both estimators put weight 1 on all observations, equally as the FD does for any d. The determination of the weights requires some tedious calculations that differ for the cases $d < 1/2$, $1/2 < d < 1$ and $d > 1$, and are left out for the ease of presentation. GLS puts higher weights on the first observations since those observations have a lower variance due to the truncation. Apparently, this leads to efficiency for $\kappa = 0$ but causes a higher impact coming from the pre-sample for $\kappa > 0$.

Figure 3-5: **: Asymptotic weights for WLS and GLS**



WLS for d=0.3, 0.7, 1, 1.3



GLS: d=0.3, 0.7, 1, 1.3

Finally, for any $0 < d < 3/2$, for a Type II with initial condition with an increasing κ (note that for $d > 1/2$, the process is not defined for $\kappa = \infty$), the asymptotic distributions of all estimators converge to the corresponding ones of a Type I process. For $d > 1/2$, this is not straightforward since the processes differ asymptotically.

In general, already for reasonably small κ, we are close to Type I behavior. In the autoregressive unit root context, such a infinite past initialization is analyzed by Phillips and Lee (1996) for the estimation of the linear trend. They show in their local to unity analysis that an increasing τ (our parameter κ) makes the GLS gain efficiency, especially for small c (their Figure 4). This contrasts our case, where the gains from efficient estimation decrease in κ.

3.4 Feasible Estimators

We have analyzed the infeasible GLS estimator relying on the knowledge of the true memory parameter d. Next, we consider the estimation with a feasible GLS where we do not know this parameter. Without an initial condition, the memory parameter can be estimated with any of the available estimation methods (e.g. parametric or semiparametric ones applied to detrended residuals, in particular the Exact Local Whittle estimator (ELW) (Shimotsu and Phillips, 2005) which is designed for Type II processes). Let $\hat{d}$ be an estimator of d with

$$(\hat{d} - d) = O_p(T^{-\tau}), \tau > 0. \tag{3.17}$$

Consequently, we substitute in equation (3.5) d by an estimator of it. Next, we show, in view of the proof of Theorem 1 in Lobato and Velasco (2007), that the feasible plug-in estimator has the same asymptotic distribution as the infeasible GLS.

Theorem 40 *For $\kappa \geq 0$, under (3.17)*

$$T^{3/2-d}\left(\hat{\beta}_{GLS(\hat{d})} - \hat{\beta}_{GLS(d)}\right) \xrightarrow{p} 0. \tag{3.18}$$

Remark 10. *There is a different strand of literature on the estimation of deterministic components under uncertainty with respect to the order of integration (see Vogelsang (1998), Bunzel and Vogelsang (2005), Harvey et al (2007) and Perron and Yabu (2009)). In particular, Vogelsang (1998) considers different estimators that are robust to I(0) and I(1) errors. Such estimators have the advantage that no estimation of the variance is needed. Two of his estimators – PS_T^2 and PSW_T^2 – can be modified to deal with fractional integration. However, in this context, the* *asymptotic distributions would depend on the memory parameter which still needs to be estimated. Thus, such an approach would lose some of its attractiveness. Harvey et al. (2010) analyze the impact of an initial condition on these robust tests and show that these tests are very sensitive to the presence of initial conditions. Thus, in our context it would be necessary to analyze the impact of our initial condition on potentially robust tests.*

We estimate the deterministic component and work with the detrended residuals (as in Shimotsu (2010)). This gives us a consistent but not efficient estimate of the memory. For the case of a Type II process without initial condition, Shimotsu (2010) proves its consistency. However, we need a memory estimate that is also robust to the presence of an initial condition and to errors that follow a Type I process. Shimotsu (2010) conjectures in his Section 4.3 that his 2-step ELW has the same asymptotic distribution under Type I errors, a result he justifies for $d < 1/2$ heuristically. In his Remark 5, he conjectures further that the presence of an initial condition of our form will have no effect on the asymptotic distribution for $d < 1/2$. However, for $d > 1/2$, it should have an effect on the asymptotic variance but not on the consistency. Own calculations comparable to the ones leading to Table 7 in Shimotsu (2010) indicate that for $1/2 < d < 1$ the Exact Local Whittle estimator appears to be consistent, even though it has a larger asymptotic variance. For $d > 1$ the estimator becomes inconsistent with $\hat{d} \to 1$ as κ increases. If we apply the Feasible ELW estimator instead, which subtracts from the series a linear combination of sample mean and first observation (see Shimotsu, 2010), the bias disappears. Consequently, we conjecture that this estimator is valid also in the presence of an initial condition. Alternatively, we might use the tapered local Whittle estimator of Velasco (1999) that is consistent in presence of trends for any d, for Type I and also for Type II (Shimotsu, 2010). Therefore, this estimator should also be consistent for the Type II process with an initial condition. Finally, if we know with certainty that $d > 1/2$, the easiest way to deal with nonstationarity and the linear trend is first differencing the data. The resulting series is stationary and the local Whittle estimator (Robinson, 1995) as well as the ELW (Shimotsu and Phillips, 2005) estimate d consistently. However if $d < 1/2$, this procedure will be inconsistent.

Dahlhaus (1995) shows in his Remark 3.2 that for Type I errors, the adaptive WLS estimator has the same variance as the infeasible one and is efficient. In practice, we estimate β with OLS or FD which are $T^{3/2-d}$-consistent, estimate d from the residuals and then reestimate β with a weighted regression (weighted mean) with weight function $w_{\hat{d}}$. From a similar argument as the one in Dahlhaus (1995), also for Type II with and without initial condition, the feasible estimator will have the same variance as the infeasible one.

The MLE depends on the knowledge of d and κ. For using an estimated d, an argument similar to the previous ones applies. For the initial condition parameter κ, we do not have any consistent estimator. Hence, we investigate numerically the effects of plugging in a wrong value of κ into the MLE (the calculated relative efficiencies are similar to previous ones and are not presented). Let κ and κ_0 denote the plugged-in and the true length of the initial condition respectively. This analysis nests the analysis of the GLS in Theorem 34, when assuming $\kappa = 0$ for $\kappa_0 > 0$. When $\kappa_0 = 0$ but the assumed $\kappa > 0$, the MLE is slightly more efficient than the WLS, especially for $d > 1$. Not surprisingly, it is still less efficient than the GLS that corresponds to the MLE that assumes correctly $\kappa = 0$. If $\kappa_0 > 0$, over- or underestimating κ in the MLE leads to virtually the same results close to the efficient ones. This implies that the MLE with a small $\kappa > 0$ outperforms the WLS when $\kappa_0 = 0$, and performs comparably to the WLS when $\kappa_0 > 0$. However, as shown in Figures 3-2 and 3-4, the gains are small.

3.5 Empirical Application: Economic Growth Rates

In order to illustrate the methodology, we apply the described estimators to the GDP growth rate series of three countries (U.S., France, Canada). Canjels and Watson (1997) estimate the annual growth rates of real GDP per capita for 128 countries. They argue that the data set is well suited since the logarithm of per-capita GDP is reasonably modeled by equation 3.1 with $(1 - \rho L)\, u_t = v_t$.

We model these series as fractional integrated processes instead. Since the paper by Gil-Alana and Robinson (1997) there has been some research on modelling the GDP alternatively as fractionally integrated (e.g. Michelacci and Zaffaroni, 2000 and Mayoral, 2006). In particular, we apply our methodology to quarterly, seasonally-adjusted series of real GDP of the U.S., France and Canada and estimate the growth rates as fractionally integrated processes. Since we have to pre-estimate d by a semiparametric estimator for the GLS and WLS, we restrict the analysis to a few countries for which the number of observations exceeds 100. For pre-estimating $\hat{d}$ we can choose any power root consistent estimator for d which can deal with trending series. We work with the feasible exact local Whittle Estimator (FELW) with prior detrending suggested by Shimotsu (2010), with a bandwidth of 0.6. This estimator accommodates an unknown mean and a polynomial trend, is consistent and has a $N(0, 1/4)$ limit distribution for the relevant region of d. We regress the log of GDP (Y_t) on $(1, t)$ and apply the FELW[4] on the residuals $\hat{Y}_t$. Table 3.1 shows the point estimates, estimated standard errors for the estimators following from

[4] See Shimotsu's homepage: http://qed.econ.queensu.ca/faculty/shimotsu/ for the code for the FELW.

Table 3.1: **Quarterly Real GDP Growth Rates in %**

Country	US	France	Canada
T	244	120	187
$\hat{\beta}_{FD}$	0.821 (0.0313)	0.5350 (0.3058)	0.876 (0.164)
CI_{FD}	[0.770; 0.878]	[0.032; 1.121]	[0.606; 1.224]
$\mathrm{CI}_{FD(\kappa=1)}$	[0.768; 0.880]	[-0.0326; 1.185]	[0.589; 1.240]
$\hat{\beta}_{OLS}$	0.824 (0.0344)	0.5264 (0.336)	0.802 (0.180)
CI_{OLS}	[0.767; 0.883]	[-0.0262; 1.170]	[0.506; 1.250]
$\mathrm{CI}_{OLS(\kappa=1)}$	[0.766; 0.884]	[-0.0755; 1.220]	[0.494; 1.262]
$\hat{\mathrm{d}}$	0.860	1.342	1.225
CI_d	[0.672;1.049]	1.1039; 1.579]	1.021; 1.429]
$\hat{\beta}_{GLS}$	0.826 (0.031)	0.618 (0.273)	0.954 (0.160)
CI_{GLS}	[0.775; 0.877]	[0.169; 1.066]	[0.691; 1.217]
$\mathrm{CI}_{GLS(\kappa=1)}$	[0.773; 0.879]	[0.0392; 1.196]	[0.668; 1.240]
$\hat{\beta}_{WLS}$	0.822 (0.031)	0.536 (0.298)	0.887 (0.163)
CI_{WLS}	[0.771; 0.873]	[0.045; 1.026]	[0.619; 1.154]
$\mathrm{CI}_{WLS(\kappa=1)}$	[0.769; 0.875]	[-0.025; 1.096]	[0.602; 1.171]

$\hat{\beta}_\vartheta$, ϑ={FD,OLS,GLS,WLS}, indicates the estimate of the FD,OLS,GLS,WLS estimator respectively with the standard errors in brackets. CI_ϑ and $\mathrm{CI}_{\vartheta(\kappa=1)}$ denote the corresponding confidence intervals under the assumption of no and the assumption of an initial condition with parameter $\kappa = 1$. $\hat{\mathrm{d}}$ denotes the memory estimate by the FELW estimator.

Theorem 34 and corresponding 5%-confidence intervals of the quarterly growth rates for the four estimators. Note that, by constructing the estimated standard error from the asymptotic result in Theorem 34, we incorporate the correlation and fractional integration structure of the errors.

For estimating the standard errors from Theorem 34, we need estimates for d and $Var(v_t)$: $\hat{d}$ and $\hat{\sigma}^2_{\hat{v}_t}$. While for the former we take the estimates using the FELW estimator, for the latter we first detrend the data and then apply the fractional difference filter to the residuals $\hat{v}_t = \Delta_t^{\hat{d}}(y_t - \hat{\mu} - \hat{\beta}t)$. Finally, we fit an AR(4) model to the residuals (chosen by the Akaike criterion) and get an approximation to $w(1)^2$ and the variance of the resulting residuals. Alternatively, we could use a nonparametric estimator for the short run dynamics.

Using any of the estimators, we reject at the 5%-level the hypothesis of *zero* growth ($\beta \leq 0$) (rows 3, 6, 12 and 15 in Table 3.1) for the U.S. and for Canada. For France, while using GLS, WLS and FD, we can reject, using the OLS we cannot reject it. The efficiency gains of using GLS and WLS are reflected by the fact that their confidence intervals are narrower.

Next, we analyze the impact of an initial condition on hypothesis testing. This allows us to answer whether we can still reject $H_0 : \beta \leq 0$ when assuming that there is a pre-sample history. From the discussion in Section 4, estimating d with this estimator provides us with a consistent estimate. From Section 3, we modify the formulae for the asymptotic variances correspondingly. When we assume $\kappa = 1$, e.g.

the formula for the variance of the GLS becomes

$$\hat{\sigma}^2_{\hat{v}_t}\left[(3-2\hat{d})\Gamma^2(2-\hat{d})+V_3\left(\hat{d},\kappa=1\right)\right]/T^{3-2\hat{d}}$$

where $V_3(\hat{d},\kappa=1)$ in (3.12) is calculated numerically for $T=10,000$.

The assumption about the initial condition increases the variance of all estimators and makes the confidence intervals wider. This effect is biggest for France ($\hat{d}=1.342$) if the GLS is adopted (as discussed in Section 3). Under this new assumption, for France even when using FD and WLS, we cannot reject the *zero* growth hypothesis anymore.

3.6 Final Remarks

We have found that the choice between the estimators depends on the assumption about presence and length of the initial condition. If we are certain about its absence, by using the GLS structure, we can obtain big efficiency gains over its alternatives. In general, the potential efficiency gains from using our GLS are, especially for $d>1$, larger than the potential losses. Thus, testing for the presence of an initial condition nesting the Type I case would be helpful. Such a test for $H_0:\kappa=0$ is left for further research.

In the unit root literature it is interesting to study the asymptotic properties of different feasible GLS estimators (cf. Canjels and Watson, 1997). In our context, the feasible GLS and WLS have the same asymptotic distribution as the infeasible ones since they only depend on the knowledge of d which can be consistently estimated. The efficient MLE, on the other hand, depends on the knowledge of κ. We find that it is only important to know whether $\kappa=0$ or $\kappa>0$, since in the former case, GLS corresponds to MLE and since in the latter case the efficiency differences are very small. Choosing a small κ in the MLE, the estimation is more robust to the initial condition than the alternative WLS, but this estimator is also more involved.

3.7 Appendix A: Lemmata

Lemma 41 *(Innovations algorithm)*
Given matrix $\sigma_u(i,j), i,j = 1:T$

$$\begin{aligned}
\hat{x}_1 &= 0 \\
\hat{x}_t &= \sum\nolimits_{i=1}^{t-1} \theta_{t-1,i}\left(u_{t-i} - \hat{x}_{t-i}\right) \text{ and} \\
v_0^* &= \sigma_u(1,1) \\
\theta_{t,t-k} &= v_k^{*-1}\left(\sigma_u(t,k) - \sum\nolimits_{j=0}^{k-1} \theta_{k,k-j}\theta_{t,t-j}v_j^*\right) \\
v_t^* &= \sigma_u(t,t) - \sum\nolimits_{j=0}^{t-1} \theta_{t-1,t-j}^2 v_j^*
\end{aligned}$$

We numerically solve the algorithm in the order $v_0^*, \theta_{11}, v_1^*, \theta_{22}, \theta_{21}, v_2^*, \theta_{33}, \theta_{32}, \theta_{31}, v_3^*$, *etc. .*

Proof. see Prop. 5.2.2 in Brockwell and Davis (1991). ■

Lemma 42 *(Terms in MLE)*

$$\begin{aligned}
&a)\ \psi_{1,1} = 1 \text{ and } \psi_{1,t} = 1 - \sum_{i=1}^{t-1} \theta_{t-1,i}\psi_{1,t-i},\ t = 2, ..., T \\
&b)\ \psi_{2,1} = 1 \text{ and } \psi_{2,t} = t - \sum_{i=1}^{t-1} \theta_{t-1,i}\psi_{2,t-i},\ t = 2, ..., T \\
&c)\ y_1^* = y_1 \text{ and } y_t^* = y_t - \sum_{i=1}^{t-1} \theta_{t-1,i}y_{t-i}^*,\ t = 2, ..., T \\
&d)\ x_1^* = x_1 \text{ and } x_t^* = x_t - \sum_{i=1}^{t-1} \theta_{t-1,i}x_{t-i}^*,\ t = 2, ..., T
\end{aligned}$$

Proof. Pattern follows from substituting recursively for $t = 1, 2$, etc. ■

Lemma 43 *(Behavior of fractional differenced regressors)*
For $0 < d < 1.5$, *for large* t,

$$\begin{aligned}
a)\ \Delta_t^d 1 &= \pi_{t-1}(d-1) \simeq \frac{(t-1)^{-d}}{\Gamma(1-d)} \\
b)\ \Delta_t^d t &= \pi_{t-1}(d-2) \simeq \frac{(t-1)^{1-d}}{\Gamma(2-d)}
\end{aligned}$$

Proof. Follows from $\Delta_t^d 1 = \sum_{k=0}^{t-1} \pi_j(d) = \pi_{t-1}(d-1) \simeq \frac{(t-1)^{-d}}{\Gamma(1-d)}$. ■

Lemma 44 *(Behavior of the GLS terms for fixed d)*
a) for $1/2 < d < 3/2$

$$\begin{aligned}
&a1) \quad q_{11}^{T} \to Q_{11}(d) = \frac{\Gamma(2d-1)}{\Gamma^2(d)} \\
&a2) \quad \textit{for } d<1,\ T^{2d-2} q_{12}^{T} \to Q_{12}(d) = \frac{1}{2\Gamma^2(2-d)} \ \textit{and} \\
&\qquad\ \textit{for } d>1,\ q_{12}^{T} \to Q_{12}(d) = O(1) \\
&a3) \quad T^{2d-3} q_{22}^{T} \to \frac{1}{(3-2d)\Gamma^2(2-d)}
\end{aligned}$$

b) for $d < 1/2$

$$\begin{aligned}
b1)\ T^{2d-1} q_{11}^{T} &\to \frac{1}{(1-2d)\,\Gamma^2(1-d)} \\
b2)\ T^{2d-2} q_{12}^{T} &\to \frac{1}{2\Gamma^2(2-d)} \\
b3)\ T^{2d-3} q_{22}^{T} &\to \frac{1}{(3-2d)\Gamma^2(2-d)}
\end{aligned}$$

Proof. a1) $q_{11}^{T} = \sum_{t=1}^{T} (\Delta_t^d 1)^2 = \sum_{t=1}^{T} \pi_{t-1}^2 (d-1) \to \frac{\Gamma(1-2(1-d))}{\Gamma^2(d)}$
where the first step follows from Lemma 3 and last step from Tanaka (1999) p.555.

a2) For $d < 1$,

$$q_{12}^{T} = \sum_{t=1}^{T} (\Delta_t^d 1)(\Delta_t^d t) \simeq 1 + \sum_{t=2}^{T} \frac{t^{-d}}{\Gamma(1-d)} \frac{t^{1-d}}{\Gamma(2-d)} \simeq 1 + \frac{T^{2-2d} - 2^{2-2d}}{2\Gamma^2(2-d)}.$$

For $d > 1$ the sum is $O(1)$ since the exponent < -1; but we cannot use the previous approximations since here the poor approximations for small t overweight.

$$\begin{aligned}
a3)\ q_{22}^{T} &= \sum_{t=1}^{T} (\Delta_t^d t)^2 \simeq \sum_{t=2}^{T} \frac{t^{2(1-d)}}{\Gamma^2(2-d)} \simeq \frac{T^{3-2d}}{(3-2d)\Gamma^2(2-d)} \\
b1)\ q_{11}^{T} &= \sum_{t=1}^{T} (\Delta_t^d 1)^2 \simeq \sum_{t=1}^{T} \frac{t^{-2d}}{\Gamma^2(1-d)} \simeq \frac{T^{1-2d}}{(1-2d)\,\Gamma^2(1-d)}
\end{aligned}$$

The proofs of b2) and b3) correspond to the ones of a2) and a3). ■

Lemma 45 *(Asymptotic distribution of the process* $u_t = \Delta_{t+[\kappa T]}^{-d} v_t$ *for* $1/2 < d < 3/2$*)*

$$T^{1/2-d} u_{[sT]} \Longrightarrow \frac{w(1)}{\Gamma(d)(2d-1)^{1/2}} W_{d-1/2}(s+\kappa) \overset{d}{=} N\left(0, w(1)^2 \frac{(s+\kappa)^{2d-1}}{\Gamma^2(d)(2d-1)}\right)$$

Proof. We first define a Type II fractional Brownian Process (see Marinucci and Robinson, 1999) as

$$W_{d-\frac{1}{2}}(t) = (2d-1)^{\frac{1}{2}} \int_0^t (t-x)^{d-1} dB(x),\ x \geq 0$$

with the covariance for $s < t$

$$E\left(W_{d-\frac{1}{2}}(s) W_{d-\frac{1}{2}}(t)\right) = \frac{1}{2}\left[s^{2d-1} + t^{2d-1} - E\left(W_{d-\frac{1}{2}}(t) - W_{d-\frac{1}{2}}(s)\right)^2\right]$$

where

$$\begin{aligned} E\left(W_{d-\frac{1}{2}}(t) - W_{d-\frac{1}{2}}(s)\right)^2 &= (2d-1)\int_0^s \left[(t-x)^{d-1} - (s-x)^{d-1}\right]^2 dx \\ &\quad +(2d-1)\int_s^t (t-x)^{2d-2} dx. \end{aligned}$$

Next, let $u_{[rT]} = \Delta_{[rT]}^{-d} \varepsilon_{[rT]}$. We know from Marinucci and Robinson (1999)[5] that

$$\Gamma(d)\left(\frac{d-\frac{1}{2}}{\pi T^{2d-1} g(0)}\right)^{\frac{1}{2}} u_{[rT]} = T^{1/2-d}\Gamma(d)(2d-1)^{\frac{1}{2}} u_{[rT]} \Rightarrow W_{d-\frac{1}{2}}(r),\ 0 < r \leq 1,$$

where we use in the first step that for $iid(0,1)$ errors the spectral density $g(0) = \frac{1}{2\pi}$. Therefore,

$$T^{1/2-d}\Delta_{[Tr]}^{-d}\varepsilon_{[Tr]} \Rightarrow \sigma(d) W_{d-\frac{1}{2}}(r),$$

where $\sigma(d) = \frac{1}{\Gamma(d)(2d-1)^{\frac{1}{2}}}$. Consequently, for $v_{[rT]} = w(L)\varepsilon_{[rT]}$, $g_v(0) = \frac{w(1)^2}{2\pi}$ and

$$T^{1/2-d}\Delta_t^{-d} v_t \Rightarrow \sigma(d) w(1) W_{d-\frac{1}{2}}(r).$$

Equally, for $u_t = \Delta_{t+[\kappa T]}^{-d} v_t$, we find that for $t = [sT]$

$$\begin{aligned} T^{1/2-d} u_{[sT]} &= T^{1/2-d} \sum_{j=0}^{[\kappa T]+[sT]} \pi_j(-d) v_{[sT]-j} \Rightarrow \sigma(d) w(1) W_{d-\frac{1}{2}}(s+\kappa) \\ &\stackrel{d}{=} N\left(0, \frac{w(1)^2 (s+\kappa)^{2d-1}}{\Gamma^2(d)(2d-1)}\right) \end{aligned}$$

since $Var\left(W_{d-\frac{1}{2}}(s+\kappa)\right) = (s+\kappa)^{2d-1}$. ■

[5] we correct a typo in their expression (3.10)

Lemma 46 *(Asymptotic distribution of the process $u_t = \Delta_{t+\kappa T}^{-d} v_t$ for $d = 1/2$)*

$$\ln T^{-1/2} u_{[sT]} = \ln T^{-1/2} \sum_{j=0}^{[\kappa T]+t} \pi_j \left(-\frac{1}{2}\right) v_{t-j} \xrightarrow{d} N\left(0, w(1)^2 \; \frac{1}{\pi}\right).$$

Hence, it does not converge to a fractional Brownian Motion and does not depend on κ.

Proof. It follows from Corollary 2.1 of Tanaka (1999). ■

Lemma 47 *(Construction of γ_j^t)*

$$\gamma_k^t = \pi_k - \sum_{j=1}^{k} \theta_{t-1,j} \gamma_{k-j}^{t-j}$$

Proof. Pattern follows from substituting recursively for $t = 1, 2$, etc. ■

3.8 Appendix B: Theorem Proofs

Proof of Theorem 34

Proof of (a)

First, the FD estimator can be rewritten as

$$T^{3/2-d}(\hat{\beta}_{FD}-\beta)=T^{3/2-d}\frac{1}{T-1}\sum\nolimits_{t=2}^{T}\Delta u_t=T^{1/2-d}(u_T-u_1).$$

To obtain the asymptotic distribution, for $0.5<d<1.5$,

$$\begin{aligned} T^{1/2-d}u_1 &= T^{1/2-d}\Delta_{1+[\kappa T]}^{-d}v_1\Rightarrow\sigma(d)w(1)W_{d-\frac{1}{2}}(\kappa) \text{ and} \\ T^{1/2-d}u_T &= T^{1/2-d}\Delta_{T+[\kappa T]}^{-d}v_T\Rightarrow\sigma(d)w(1)W_{d-\frac{1}{2}}(1+\kappa), \end{aligned}$$

from Lemma 45(for $s=0$ and 1). Next,

$$Var\left(T^{1/2-d}\left(u_T-u_1\right)\right)=\frac{(\kappa+1)^{2d-1}+(\kappa)^{2d-1}}{\Gamma^2(d)(2d-1)}-2V_1\left(d,\kappa\right)$$

where

$$\begin{aligned} V_1\left(d,\kappa\right) &= Cov(T^{1/2-d}u_T,T^{1/2-d}u_1)=T^{1-2d}E\left(\sum_{j=0}^{[\kappa T]}\pi_j(-d)\pi_{T+j}(-d)v_{t-j}^2\right) \\ &= T^{1-2d}\sum_{j=0}^{[\kappa T]}\pi_j(-d)\pi_{T+j}(-d)\simeq\frac{T^{1-2d}\sum\limits_{j=0}^{[\kappa T]}j^{d-1}\left(T+j\right)^{d-1}}{\Gamma^2(d)} \\ &\simeq\frac{\int_0^{\kappa}\left(s+s^2\right)^{d-1}ds}{\Gamma^2(d)}. \end{aligned}$$

Proof of (b)

Write

$$\hat{\beta}_{OLS}-\beta=\frac{\frac{1}{T}\sum\limits_{t=1}^{T}tu_t-\left(\frac{1}{T}\sum\limits_{t=1}^{T}t\right)\left(\frac{1}{T}\sum\limits_{t=1}^{T}u_t\right)}{\frac{1}{T}\sum\limits_{t=1}^{T}t^2-\left(\frac{1}{T}\sum\limits_{t=1}^{T}t\right)^2}. \tag{3.19}$$

For the denominator,

$$\frac{1}{T^2}\left[\frac{1}{T}\sum_{t=2}^{T}t^2-\left(\frac{1}{T}\sum_{t=2}^{T}t\right)^2\right]\to\frac{1}{12}.$$

Using Lemma 45, we obtain that

$$T^{-1}\sum_{t=2}^{T} T^{1/2-d}u_t \Rightarrow \frac{w(1)}{\Gamma(d)(2d-1)^{\frac{1}{2}}}\int_0^1 W_{d-\frac{1}{2}}(\kappa+s)ds$$

and that

$$\frac{1}{T}\sum_{t=2}^{T}\left(\frac{t}{T}-\frac{1}{2}\right)T^{1/2-d}u_t \Rightarrow \frac{w(1)}{\Gamma(d)(2d-1)^{\frac{1}{2}}}\left\{\int_0^1\left(s-\frac{1}{2}\right)W_{d-\frac{1}{2}}(\kappa+s)ds\right\}.$$

Since it is a function of fractional Brownian Motion it is clearly Gaussian. For the variance,

$$\begin{aligned}
& Var\left(\int_0^1\left(s-\frac{1}{2}\right)W_{d-\frac{1}{2}}(\kappa+s)ds\right) \\
= & \; Cov\left(\int_0^1\left(s-\frac{1}{2}\right)W_{d-\frac{1}{2}}(\kappa+s)ds, \int_0^1\left(t-\frac{1}{2}\right)W_{d-\frac{1}{2}}(\kappa+t)dt\right) \\
= & \int_0^1\int_0^1\left(s-\frac{1}{2}\right)\left(t-\frac{1}{2}\right)E\left\{W_{d-\frac{1}{2}}(\kappa+s)W_{d-\frac{1}{2}}(\kappa+t)\right\}dsdt \\
= & \; 2\int_0^1\left\{\int_0^t\left(s-\frac{1}{2}\right)\left(t-\frac{1}{2}\right)\frac{1}{2}\left[(\kappa+s)^{2d-1}+(\kappa+t)^{2d-1}\right.\right. \\
& -(2d-1)\int_0^s\{(\kappa+t-x)^{d-1}-(\kappa+s-x)^{d-1}\}^2dx \\
& \left.-(2d-1)\int_s^t(\kappa+t-x)^{2d-2}dx\right]dsdt
\end{aligned}$$

which equals to (3.11). All steps in this proof are valid for $0.5<d<1.5$. Notice that for $d=1$, the fractional BM becomes a standard BM.

Proof of (c)

The GLS estimator can be written as

$$\hat{\beta}_{GLS}=\beta+\frac{(q_{11}^T r_2^T-q_{12}^T r_1^T)}{(q_{11}^T q_{22}^T-(q_{12}^T)^2)} \tag{3.20}$$

where q_{11} and q_{12} are defined in (3.7) and

$$r_1^T=\sum_{t=1}^{T}(\Delta_t^d 1)v_t+\sum_{t=1}^{T}(\Delta_t^d 1)\left(\Delta_t^d\xi_t\right) \quad \text{and} \quad r_2^T=\sum_{t=1}^{T}(\Delta_t^d t)v_t+\sum_{t=1}^{T}(\Delta_t^d t)\left(\Delta_t^d\xi_t\right).$$

Using Lemma 44, we obtain for the denominator

$$\frac{1}{T^{3-2d}}(q_{11}^T q_{22}^T-\left(q_{12}^T\right)^2)\rightarrow\frac{\Gamma(2d-1)}{\Gamma^2(d)}\frac{1}{(3-2d)\Gamma^2(2-d)} \tag{3.21}$$

since for $d > 0.5$, $\frac{T^{4-4d}}{T^{3-2d}} = T^{1-2d} = o(1)$ and consequently the order of the second term is negligible.

The numerator consists of two uncorrelated terms

$$\begin{aligned} q_{11}^T r_2^T - q_{12}^T r_1^T &= \sum_{t=1}^{T} \chi_t^T(d) \left(v_t + \sum_{k=0}^{t-1} \pi_k(d) \sum_{j=0}^{[\kappa T]} \pi_{t-k+j}(-d) v_j \right) \\ &= ... = \sum_{t=1}^{T} \chi_t^T(d) v_t + \sum_{j=0}^{[\kappa T]} \left(\sum_{t=1}^{T} \chi_t^T \sum_{k=0}^{t-1} \pi_k(d) \pi_{t-k+j}(-d) \right) v_{-j} \quad (3.22) \end{aligned}$$

where $\chi_t(d)$ is defined as (3.14). For the first term of (3.22), note that

$$r_1^T = \sum_{t=1}^{T} (\Delta_t^d 1) v_t = O_p(1)$$

since by Lemma 44

$$Var(r_1^T) = Var\left(\sum_{t=1}^{T} (\Delta_t^d 1) w(L) \varepsilon_t \right) = w(1)^2 \frac{\Gamma(2d-1)}{\Gamma^2(d)}.$$

This implies that

$$\frac{1}{T^{3/2-d}} q_{12}^T r_1^T \xrightarrow{d} 0 \quad (3.23)$$

since $\frac{T^{2-2d}}{T^{3/2-d}} = T^{1/2-d} = o(1)$ for $d > 0.5$. Next, $r_2^T = \sum_{t=1}^{T} (\Delta_t^d t) v_t$ with

$$Var\left(\frac{1}{T^{3/2-d}} r_2^T \right) \xrightarrow{p} \frac{w(1)^2}{\Gamma^2(2-d)(3-2d)}$$

by Lemma 46, implying that

$$\frac{1}{T^{3/2-d}} q_{11}^T r_2^T \xrightarrow{d} N\left(0, \left(\frac{\Gamma(2d-1)}{\Gamma^2(d)} \right)^2 \frac{w(1)^2}{\Gamma^2(2-d)(3-2d)} \right) \quad (3.24)$$

Combining, we obtain for the first term of the numerator

$$\frac{1}{T^{3/2-d}} (q_{11}^T r_2^T - q_{12}^T r_1^T) \xrightarrow{d} N\left(0, \left(\frac{\Gamma(2d-1)}{\Gamma^2(d)} \right)^2 \frac{w(1)^2}{\Gamma^2(2-d)(3-2d)} \right) \quad (3.25)$$

where the covariance of the two terms is of lower order so that it goes to zero. The second term of (3.22) adds to the variance the term (3.13) which needs to be calculated numerically. Finally, combining denominator and numerator gives the result. For $1 < d < 1.5$, all steps remain the same except that from Lemma 44

$$q_{12}^T - 1 \rightarrow Q_{12}(d) = O(1)$$

This has no effect on the asymptotic distribution. Finally, for $d = 1$, the GLS corresponds to the FD estimator.

Proof of (d)

The WLS can be written as

$$\hat{\beta}_{GLS} - \beta = \frac{\sum_{t=1}^{T} \left(\frac{t}{T+1}\right)^{1-d} \left(1 - \frac{t}{T+1}\right)^{1-d} \Delta u_i}{\sum_{t=1}^{T} \left(\frac{t}{T+1}\right)^{1-d} \left(1 - \frac{t}{T+1}\right)^{1-d}}. \tag{3.26}$$

The denominator of (3.26) behaves asymptotically as

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \left(\frac{t}{T+1}\right)^{1-d} \left(1 - \frac{t}{T+1}\right)^{1-d} &\simeq \int_0^1 x^{1-d}\left(1-x\right)^{1-d} dx \\ &= \frac{\Gamma^2\left(2-d\right)}{\Gamma\left(4-2d\right)} = \frac{\Gamma^2\left(2-d\right)}{\Gamma\left(3-2d\right)\left(3-2d\right)}. \end{aligned} \tag{3.27}$$

For the numerator of (3.26), we find

$$\begin{aligned} & T^{1/2-d}\sum_{t=1}^{T} w_d\left(\frac{t}{T+1}\right)\Delta u_t \\ = \;& T^{1/2-d}\sum_{t=1}^{T} w_d\left(\frac{t}{T+1}\right)\left[\sum_{k=0}^{t-1}\pi_k\left(1-d\right)v_{t-k} + \sum_{k=0}^{[\kappa T]-1}\pi_{t+k}\left(1-d\right)v_{-k}\right] \end{aligned} \tag{3.28}$$

where we use that

$$\begin{aligned} \Delta u_t &= \Delta\left(\sum_{j=1}^{t}\sum_{k=0}^{j-1}\pi_k\left(-d+1\right)v_{j-k} + \sum_{k=0}^{[\kappa T]-1}\sum_{j=0}^{t+k}\pi_j\left(-d+1\right)v_{-k}\right) \\ &= \sum_{k=0}^{t-1}\pi_k\left(-d+1\right)v_{t-k} + \sum_{k=0}^{[\kappa T]-1}\Delta\sum_{j=0}^{t+k}\pi_j\left(-d+1\right)v_{-k}. \end{aligned}$$

Since both terms in (3.28) are uncorrelated, the variance is the sum of the respective variances. For the first term of (3.28),

$$T^{1/2-d}\sum_{t=1}^{T} w_d\left(\frac{t}{T+1}\right)\Delta u_t = T^{1/2-d}\sum_{t=1}^{T}\left(\sum_{k=0}^{T-t}\pi_k\left(1-d\right)w_{t+k}\right)v_t \tag{3.29}$$

since

$$
\begin{aligned}
& T^{1/2-d}\sum_{t=1}^{T} w_d\left(\frac{t}{T+1}\right)\Delta u_t = T^{1/2-d}\sum_{t=1}^{T}\left(\frac{t}{T+1}\right)^{-d}\left(1-\frac{t}{T+1}\right)^{-d}\Delta_t^{1-d}v_t \\
& = T^{1/2-d}\sum_{t=1}^{T} w_d\left(\frac{t}{T+1}\right)\Delta_t^{1-d}v_t = T^{1/2-d}\sum_{t=1}^{T} w_d\left(\frac{t}{T+1}\right)\left(\sum_{k=0}^{t-1}\pi_k\left(1-d\right)v_{t-k}\right) \\
= \quad & T^{1/2-d}\sum_{t=1}^{T}\left(\sum_{k=0}^{T-t}\pi_k\left(1-d\right)w_d\left(\frac{t+k}{T+1}\right)\right)v_t.
\end{aligned}
$$

The corresponding variance

$$
T^{1-2d}\sum_{t=1}^{T}\left(\sum_{k=0}^{T-t}\pi_k\left(1-d\right)w_d\left(\frac{t+k}{T+1}\right)\right)^2
$$

converges to $w^2\left(1\right)V_{41}\left(d\right)$ in (3.15). The second term of (3.28) becomes

$$
\begin{aligned}
& \sum_{k=0}^{[\kappa T]-1}\sum_{j=0}^{t+k}\pi_j\left(-d+1\right)v_{-k} - \sum_{k=0}^{[\kappa T]-1}\sum_{j=0}^{t-1+k}\pi_j\left(-d+1\right)v_{-k} \\
= \quad & \sum_{k=0}^{[\kappa T]-1}\pi_{t+k}\left(-d+1\right)v_{-k}.
\end{aligned}
$$

Thus, the corresponding variance is

$$
\begin{aligned}
& Var\left[T^{1/2-d}\sum_{t=1}^{T} w_d\left(\frac{t}{T+1}\right)\sum_{k=0}^{[\kappa T]-1}\pi_{t+k}\left(1-d\right)v_{-k}\right] \\
= \quad & T^{1-2d}Var\left[\sum_{k=0}^{[\kappa T]-1}\left(\sum_t w_d\left(\frac{t}{T+1}\right)\pi_{t+k}\left(1-d\right)\right)v_{-k}\right] \\
= \quad & w\left(1\right)^2 T^{1-2d}\sum_{k=0}^{[\kappa T]-1}\left(\sum_t w_d\left(\frac{t}{T+1}\right)\pi_{t+k}\left(1-d\right)\right)^2
\end{aligned}
$$

Next, we approximate this term by

$$
\begin{aligned}
& T^{1-2d}T\frac{1}{T}\sum_{k=0}^{[\kappa T]-1}\left(T\frac{1}{T}\sum\nolimits_{t=1}^{T}\left(\frac{t}{T+1}\right)^{1-d}\left(1-\frac{t}{T+1}\right)^{1-d}\frac{T^{d-2}\left(\frac{t+k}{T}\right)^{d-1-1}}{\Gamma\left(d-1\right)}\right)^2 \\
= \quad & \frac{1}{T}\sum_{k=0}^{[\kappa T]-1}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T+1}\right)^{1-d}\left(1-\frac{t}{T+1}\right)^{1-d}\frac{\left(\frac{t+k}{T}\right)^{d-1-1}}{\Gamma\left(d-1\right)}\right)^2,
\end{aligned}
$$

converging to $w\left(1\right)^2 V_{42}\left(d,\kappa\right)$ in (3.16).

Proof of (e)

The MLE can be written as

$$\hat{\beta}_{MLE} = \beta + \frac{\left(q_{11}^T r_2^T - q_{12}^T r_1^T\right)}{\left(q_{11}^T q_{22}^T - \left(q_{12}^T\right)^2\right)}$$

where q_{11} and q_{12} are defined in (3.10) and

$$r_1^T = \sum_{t=1}^{T} \psi_{1t} x_t^* / v_t^* \text{ and } r_2^T = \sum_{t=1}^{T} \psi_{2t} x_t^* / v_t^*$$

First, for $\kappa = 0$, it follows from the construction of the weights in the BLP that $\theta_{.,i} = \pi_i(-d)$. As a consequence, the filter cancels the fractional integration, leading to $x_t^* = v_t$ (since $x_1^* = x_1 = v_1$ and $x_2^* = x_2 - \theta_{1,1} x_1^* = v_2 + \pi_1(-d) v_1 - \theta_{1,1} v_1 = v_2$ etc.). This implies that the variance of the estimator is:

$$T^{3-2d} \frac{Var\left[\sum_{t=1}^{T} \left(q_{11}^T \psi_{2t}/\sqrt{v_t^*} - q_{12}^T \psi_{1t}/\sqrt{v_t^*}\right) x_t^*/\sqrt{v_t^*}\right]}{\left(q_{11}^T q_{22}^T - \left(q_{12}^T\right)^2\right)^2} \tag{3.30}$$

$$= T^{3-2d} \frac{\sum_{t=1}^{T} \left(q_{11}^T \psi_{2t} - q_{12}^T \psi_{1t}\right)^2}{\left(q_{11}^T q_{22}^T - \left(q_{12}^T\right)^2\right)^2}$$

Finally, note that ψ_{1t} and ψ_{2t} correspond to $\pi_t(d-1)$ and $\pi_t(d-2)$ respectively, implying that the MLE corresponds algebraically to the GLS. For $\kappa > 0$, ψ_{1t} and ψ_{2t} depend on κ and are defined in Lemma 42. $v_t^*, \psi_{1t}, \psi_{2t}$ are now different than the corresponding terms in the GLS estimator and are generated in the Lemmata 41 and 42. Further note that in this case, x_t^* can be written as

$$x_t^* = \sum_{j=0}^{t+[\kappa T]} \gamma_j^t v_{t-j},$$

where the weights γ_j^t are constructed following Lemma 47. Nevertheless, due to its construction in the Innovations Algorithm, the resulting forecast errors are uncorrelated and we obtain for the variance of the MLE again expression (3.30) with correspondingly defined terms ψ_{1t} and ψ_{2t}.

Proof of Theorem 36

Proof of c)

For the numerator of the GLS, we need $Var\left(\sum_{t=1}^{T} \chi_{t-1}^T(d) \Delta_t^d u_t\right)$ where u_t is the nonstationary Type I process (3.3) and

$$\chi_{t-1}^T(d) = q_{11}^T \pi_{t-1}(d-2) - q_{12}^T \pi_{t-1}(d-1) \tag{3.31}$$

with q_{11} and q_{12} defined in the Proof of Theorem 1. It can be shown that it equals to

$$\begin{aligned} & Var\left(\sum_{t=1}^{T}\left(\sum_{k=0}^{T-t}\chi_{t-1}^{T}\left(d\right)\pi_{k}\left(d\right)\right)u_{t}\right) \\ = & \sum_{t=1}^{T}\sum_{s=1}^{T}\left(\sum_{k=0}^{T-t}\chi_{t-1}^{T}\left(d\right)\pi_{k}\left(d\right)\right)\left(\sum_{l=0}^{T-s}\chi_{t-1}^{T}\left(d\right)\pi_{l}\left(d\right)\right)Cov\left(u_{s},u_{t}\right). \end{aligned} \tag{3.32}$$

As shown in Marinucci and Robinson (1999), the process (3.3) multiplied by T^{1-2d} converges to a fractional BM of Type I with covariance

$$\frac{\Gamma(3-2d)}{(2d-1)\Gamma(d)\Gamma(2-d)}\frac{1}{2}\left(s^{2d-1}+t^{2d-1}-|t-s|^{2d-1}\right)$$

For finite T, the approximate covariance for $1 \leq s \leq t \leq T$ becomes

$$Cov\left(u_{t},u_{s}\right)=\frac{\Gamma(3-2d)}{(2d-1)\Gamma(d)\Gamma(2-d)}\frac{1}{2}\left(s^{2d-1}+t^{2d-1}-|t-s|^{2d-1}\right). \tag{3.33}$$

Substituting the covariance in (3.32) gives the result.

Proof of e)

The MLE for Type I corresponds to the one for Type II with a different covariance matrix $\sigma_u(i,j)$. From the discussion of Proof d) it follows that $\sigma_u(i,j) = Cov(x_i, x_j)$ given in (3.33).

Proof of Proposition 37

Proof of (a)

First,

$$\begin{aligned} & T\left(\ln T\right)^{-1/2}\left(\hat{\beta}_{FD}-\beta\right)=\left(\ln T\right)^{-1/2}\left(\Delta_{T+[\kappa T]}^{-1/2}v_{T}-\Delta_{1+[\kappa T]}^{-1/2}v_{1}\right) \\ = & \left(\ln T\right)^{-1/2}\left(\sum_{j=0}^{T-1}\pi_{j}\left(-\frac{1}{2}\right)v_{T-j}+\sum_{j=0}^{[\kappa T]}\left(\pi_{T+j}\left(-\frac{1}{2}\right)-\pi_{j}\left(-\frac{1}{2}\right)\right)v_{-j}\right) \end{aligned}$$

has zero mean and consists of two uncorrelated terms leading to a variance

$$\left(\ln T\right)^{-1}\sum_{j=0}^{T-1}\pi_{j}\left(-\frac{1}{2}\right)^{2}+\left(\ln T\right)^{-1}\sum_{j=0}^{[\kappa T]}\left(\pi_{T+j}\left(-\frac{1}{2}\right)-\pi_{j}\left(-\frac{1}{2}\right)\right)^{2}.$$

The first term converges to $1/\pi$. The second term is of order $o_p(1)$ for $\kappa = 0$ and converges to $1/\pi$ for any $\kappa > 0$. This follows from

$$
\begin{aligned}
& (lnT)^{-1} \sum_{j=0}^{[\kappa T]} \left(\pi_{T+j} \left(-\frac{1}{2} \right) - \pi_j \left(-\frac{1}{2} \right) \right)^2 \\
\simeq \quad & \frac{1}{\pi} (lnT)^{-1} \sum_{j=0}^{[\kappa T]} \left\{ \frac{(T+j)^{-1}}{\ln T} - 2 \frac{(T+j)^{-1/2} j^{-1/2}}{\ln T} + \frac{(j)^{-1}}{\ln T} \right\},
\end{aligned}
$$

where the last term converges to $1/\pi$ and the first two terms vanish since

$$
\sum_{j=0}^{[\kappa T]} \frac{(T+j)^{-1}}{\ln T} = \sum_{j=T}^{T+[\kappa T]} \frac{j^{-1}}{\ln T} = \frac{\ln (T + \kappa T)}{\ln T} - \frac{\ln T}{\ln T} = \frac{\ln T}{\ln T} + \frac{\ln (1+\kappa)}{\ln T} - \frac{\ln T}{\ln T} \to 0.
$$

Proof of Theorem 38

Proof of a)

$T \left(\hat{\beta}_{FD} - \beta \right)$ has mean zero and a variance

$$
Var(u_T) + Var(u_1) - 2Cov\left(u_T, u_1\right).
$$

The first term converges to $\sum_{j=0}^{\infty} \pi_j \left(-d\right)^2 = \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$, where the last step follows from the proof of Lemma 4a1). The covariance converges to zero from a similar argument to the one in Proposition 37. Finally, for $\kappa = 0$, $Var(u_1) = 1$ and for $\kappa > 0$, the term converges equally to $\frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$. Since no central limit theorem is applicable, the asymptotic distribution depends on the distribution of the innovations ε_t.

Proof of b)

First, note that the CLT for fractional integrated processes are not valid here. As a consequence, we analyze the limit distribution of the numerator of (3.19) directly. Write

$$
u_t = \sum_{j=0}^{[\kappa T]+t-1} \pi_j \left(-d\right) v_{t-j} = \Delta_t^{-1/2} v_t + \sum_{j=t}^{[\kappa T]+t-1} \pi_j \left(-d\right) v_{t-j}.
$$

The first term

$$
\begin{aligned}
T^{-1-2d} \sum_{t=2}^{T} \left(\frac{t}{T} - \frac{1}{2} \right) \Delta_t^{-d} v_t \quad &= \quad T^{-1-2d} \sum_{t=2}^{T} \left(\frac{t}{T} - \frac{1}{2} \right) \sum_{j=0}^{t-1} \pi_j \left(-d\right) v_{t-j} = ... = \\
&= \quad T^{-1-2d} \sum_{k=1}^{T} \left[\sum_{j=0}^{T-k} \left(\frac{j+k}{T} - \frac{1}{2} \right) \pi_j \left(-d\right) \right] v_k,
\end{aligned}
$$

leading for $v_t = \varepsilon_t$ to a variance

$$T^{-1-2d} \sum\nolimits_{k=1}^{T} \left(\sum\nolimits_{j=0}^{T-k} \left(\frac{j+k}{T} - \frac{1}{2} \right) \pi_j \left(-d\right) \right)^2,$$

which we can approximate through Riemann sums by

$$\int_0^1 \left(\int_0^{1-k} \left(s + k - \frac{1}{2} \right) \frac{s^{d-1}}{\Gamma\left(d\right)} ds \right)^2 dk = \frac{\left(2d^3 - d^2 + 1\right)}{4\Gamma^2\left(2+d\right)\left(4d^2 + 8d + 3\right)}.$$

For the second term note that

$$\sum_{t=2}^{T} \left(\frac{t}{T} - \frac{1}{2} \right) \sum_{j=t}^{[\kappa T]+t} \pi_j \left(-d\right) v_{t-j} = ... = \sum_{k=0}^{[\kappa T]} \left[\sum_{j=2}^{T} \left(\frac{j}{T} - \frac{1}{2} \right) \pi_{k+j} \left(-d\right) \right] v_{-k}.$$

Thus, by approximating this term again through Riemann sums, this term adds

$$\int_0^{\kappa} \left(\int_0^1 \left(s - \frac{1}{2} \right) \frac{\left(k+s\right)^{d-1}}{\Gamma\left(d\right)} ds \right)^2 dk.$$

to the variance. Combining with the denominator, which converges to 1/12, we obtain the result.

Proof of c)

From Lemmata 3 and 4, we find that both terms in the denominator are of the same order implying that

$$\frac{1}{T^{4-4d}} (q_{11}^T q_{22}^T - \left(q_{12}^T\right)^2) \rightarrow \frac{1}{\left(1-2d\right)\Gamma^2(1-d)} \frac{1}{(3-2d)\Gamma^2(2-d)} - \frac{1}{4\Gamma^4(2-d)}. \tag{3.34}$$

The numerator (3.22) consists of two terms. From a similar argument as the one in the Proof of Theorem 34c), now when multiplied by $T^{3d-5/2}$, both terms in the numerator are of the same order. Thus, we obtain from Lemmata 43 and 44 for $0 < d < 1/2$,

$$\begin{aligned}
&Var\left(\frac{1}{T^{1-2d}} q_{11}^T \frac{1}{T^{3/2-d}} r_2^T \right) \xrightarrow{p} \frac{1}{\left(1-2d\right)^2 \Gamma^4(1-d)} \frac{1}{(3-2d)\Gamma^2(2-d)}, \\
&Var\left(\frac{1}{T^{2-2d}} q_{12}^T \frac{1}{T^{1/2-d}} r_1^T \right) \xrightarrow{p} \frac{1}{4\Gamma^4(2-d)} \frac{1}{\left(1-2d\right)\Gamma^2(1-d)}, \text{ and} \\
&Cov\left(\frac{1}{T^{5/2-3d}} \; q_{11}^T r_2^T, \frac{1}{T^{5/2-3d}} q_{12}^T r_1^T \right) \xrightarrow{p} \frac{1}{\left(1-2d\right)\Gamma^2(1-d)} \frac{1}{2\Gamma^2(2-d)} \frac{1}{2\Gamma^2(2-d)}.
\end{aligned}$$

Hence, we find for the first term of the numerator for $v_t = w(L)\varepsilon_t$,

$$\frac{1}{T^{5/2-3d}}(q_{11}^T r_2^T - q_{12}^T r_1^T) \xrightarrow{d} N\left(0, \frac{w(1)^2}{4\Gamma^4(1-d)\Gamma^2(2-d)(3-2d)(2d^2-3d+1)^2}\right) \quad (3.35)$$

The second term of (3.22) adds to the variance the term $V_{82}(d,\kappa)$. Finally, combining the numerator and the denominator gives the result.

Proof of d)

Corresponds to the proof of Theorem 34 d).

Proof of Theorem 39.

Proof of b)

The argument corresponds to the one of Theorem 36 b).

Proof of c)

The stationary process (3.2) has a covariance

$$Cov(u_i, u_j) = \gamma(|i-j|) = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \frac{\Gamma(|i-j|+d)}{\Gamma(1+|i-j|-d)}. \quad (3.36)$$

(see e.g. Hosking, 1996). Substituting $Cov(u_i, u_j)$ in (3.32) gives the variance of the GLS estimator.

Proof of d)

See Dahlhaus (1995).

Proof of e)

Parallel to the discussion in proof of Theorem 36, we obtain the MLE by noting that $\kappa(i,j) = Cov(x_i, x_j)$ given in (3.36).

Proof of Theorem 40

First, we assume $\kappa = 0$. In view of the proof in Lobato and Velasco (2007), we can show that

$$T^{3/2-d}\left(\hat{\beta}_{GLS(\hat{d})} - \hat{\beta}_{GLS(d)}\right) \xrightarrow{p} 0,$$

where we analyze only the most critical component. This is the scaled numerator in the asymptotic distribution. For $1/2 < d < 3/2$, the dominating term is $q_{11}(d)r_2(d)$. Hence, we have to show that

$$\frac{1}{T^{3/2-d}}\left(q_{11}(d)r_2(d) - q_{11}(\hat{d})r_2(\hat{d})\right) \xrightarrow{p} 0.$$

The term equals

$$\frac{1}{T^{3/2-d}}\left[q_{11}(d)\left(\sum_{t=1}^{T}(\Delta_t^d t)v_t\right) - q_{11}\left(\hat{d}\right)\left(\sum_{t=1}^{T}(\Delta_t^{\hat{d}} t)v_t\right)\right]$$

$$\simeq \frac{1}{T^{3/2-d}}\sum_{t=1}^{T}\underbrace{\left[Q_{11}(d)(\Delta_t^d t) - Q_{11}(\hat{d})(\Delta_t^{\hat{d}} t)\right]}_{S_t(d)} v_t$$

This mean *zero* process has a variance of $\frac{1}{T^{3-2d}}\sum_{t=1}^{T}S_t(d)^2w(1)^2$. Note that for $t = 2,3,...,T$,

$$S_t(d)=\sum_{r=1}^{R-1}\frac{1}{r!}\left(d-\hat{d}\right)^r\frac{\partial^r S_t(d)}{\partial d^r}+\frac{1}{R!}\left(d-\hat{d}\right)^R\frac{\partial^R S_t(\bar{d})}{\partial d^R} \tag{3.37}$$

where $\bar{d}$ is an intermediate point between d and $\hat{d}$. Note that

- $\left(d-\hat{d}\right)=o_p(T^{-\tau})$,
- $Q_{11}(d)=\frac{\Gamma(2d-1)}{\Gamma^2(d)}=O(1)$ and $Q_{11}(d)^{(1)}=\frac{\Gamma^2(d)\dot{\Gamma}(2d-1)2-\Gamma(2d-1)2\Gamma(d)\dot{\Gamma}(d)}{\Gamma^4(d)}=O(1)$ implying that $Q_{11}(d)^{(r)}=O(1)$ and
- $(\Delta_t^d t)=\pi_{t-1}\left(d-2\right)\simeq\frac{t^{1-d}}{\Gamma(2-d)}$ and $(\Delta_t^d t)^{(r)}=|\pi_{t-1}^{(r)}\left(d-2\right)|$
  $\leq C\left(t-1\right)^{-d+2-1}\log^r\left(t-1\right)=o\left(t^{1-d+\varepsilon}\right)$

where the inequality follows from Wright (1995). Consequently, for t=2,...,T

$$\begin{aligned}
\frac{\partial S_t(d)}{\partial d} &= Q_{11}(d)^{(1)}(\Delta_t^d t)+Q_{11}(d)\left(\Delta_t^d t\right)^{(1)}=o\left(t^{1-d+\varepsilon}\right), \\
\frac{\partial^2 S_t(d)}{\partial d^2} &= Q_{11}(d)^{(2)}(\Delta_t^d t)+2Q_{11}(d)^{(1)}\left(\Delta_t^d t\right)^{(1)}+Q_{11}(d)\left(\Delta_t^d t\right)^{(2)}=o\left(t^{1-d+\varepsilon}\right), \qquad (3.38)\\
\frac{\partial^r S_t(d)}{\partial d^r} &= o\left(t^{1-d+\varepsilon}\right).
\end{aligned}$$

Using (3.38), (3.37) can be written as

$$S_t(d)\simeq\left(d-\hat{d}\right)o\left(t^{1-d+\varepsilon}\right)+o(t^{1-d}).$$

This implies that

$$\frac{1}{T^{3-2d}}\sum_{t=1}^{T}S(d)^2w(1)^2=\frac{1}{T^{3-2d}}\sum_{t=1}^{T}\left(o_p(T^{-\tau})o\left(t^{1-d+\varepsilon}\right)\right)^2+o\left(1\right)=o_p(T^{-2\tau+\varepsilon})\xrightarrow{p}0$$

and establishes the result.

Following the same lines, we can show for $d<1/2$ that

$$\frac{1}{T^{3/2-d}}\left(q_{11}(d)r_2(d)-q_{12}(d)r_1(d)-\left(q_{11}(\hat{d})r_2(\hat{d})-q_{12}(\hat{d})r_1(\hat{d})\right)\right)\xrightarrow{p}0.$$

For $\kappa>0$, in addition to the analyzed term there are also some terms coming from the initial condition. Consequently, we have to show that

$$\frac{1}{T^{3/2-d}}\left((q_{11}(d)r_2(d)-q_{12}(d)r_1(d))-\left(q_{11}(\hat{d})r_2(\hat{d})-q_{12}(\hat{d})r_1(\hat{d})\right)\right)\xrightarrow{p}0.$$

In particular, we show that, for $1/2 < d < 3/2$,

$$Var\left(\frac{1}{T^{3/2-d}}\sum_{j=0}^{[\kappa T]-1}\left(\sum_{t=1}^{T}S_{t,j}\left(d\right)\right)v_{-j}\right)=\frac{1}{T^{3-2d}}\sum_{j=0}^{[\kappa T]-1}\left(\sum_{t=1}^{T}S_{t,j}\left(d\right)\right)^{2}=o\left(1\right),$$

where

$$S_{t,j}\left(d\right)=\left[\left(\psi_{t}\left(d\right)\sum_{k=0}^{t-1}\pi_{k}\left(d\right)\pi_{t-k+j}\left(-d\right)\right)-\left(\psi_{t}\left(\hat{d}\right)\sum_{k=0}^{t-1}\pi_{k}\left(\hat{d}\right)\pi_{t-k+j}\left(-\hat{d}\right)\right)\right].$$

For the Taylor approximation of $S_{t,j}\left(d\right)$ similar to (3.37), we need

$$\frac{\partial S_{t,j}}{\partial d}=\dot{\chi}_{t}^{T}\left(d\right)\sum_{k=0}^{t-1}\pi_{k}\left(d\right)\pi_{t-k+j}\left(-d\right)+\chi_{t}^{T}\left(d\right)\frac{\partial}{\partial d}\left(\sum_{k=0}^{t-1}\pi_{k}\left(d\right)\pi_{t-k+j}\left(-d\right)\right)$$

where the different terms behave as follows:

- $\chi_{t}^{T}\left(d\right)\simeq\left(q_{11}^{T}\pi_{t-1}\left(d-2\right)-q_{12}^{T}\pi_{t-1}\left(d-1\right)\right)\simeq O\left(t^{1-d}\right)$
- $\dot{\chi}_{t}^{T}\left(d\right)\simeq\left(\dot{q}_{11}^{T}\pi_{t-1}\left(d-2\right)+q_{11}^{T}\dot{\pi}_{t-1}\left(d-2\right)-\dot{q}_{12}^{T}\pi_{t-1}\left(d-1\right)-q_{12}^{T}\dot{\pi}_{t-1}\left(d-1\right)\right)\simeq o\left(t^{1-d+\varepsilon}\right)$
- $\sum_{k=0}^{t-1}\pi_{k}\left(d\right)\pi_{t-k+j}\left(-d\right)=K_{1}\sum_{k=0}^{t-2}\left(t-k+j\right)^{d-2}k^{-d}+K_{2}\left(j+1\right)^{d-1}t^{-d}\leq K_{3}t^{-1}+K_{4}t^{-d}j^{d-1}$, where the first step follows from summation by partsand in the second step we use: $\sum_{k=0}^{t-2}\left(t-k+j\right)^{d-2}k^{-d}\leq\sum_{k=0}^{t/2-1}\left(\frac{t}{2}+j\right)^{d-2}k^{-d}+\sum_{k=t/2}^{t-2}\left(t-k+j\right)^{d-2}\left(\frac{t}{2}\right)^{-d}$ $=\left(\frac{t}{2}+j\right)^{d-2}t^{1-d}+\left(\frac{t}{2}\right)^{-d}\sum_{k=j}^{t/2+j}k^{d-2}\simeq\left(\frac{t}{2}+j\right)^{d-2}t^{1-d}+\left(\frac{t}{2}\right)^{-d}\frac{(t/2+j)^{d-1}-j^{d-1}}{d-1}$, which for $d<1$ is bounded by $K_{5}t^{-1}+K_{6}t^{-d}j^{d-1}$ and for $d>1$ by $K_{5}t^{-1}$.
- $\frac{\partial}{\partial d}\sum_{k=0}^{t-1}\pi_{k}\left(d\right)\pi_{t-k+j}\left(-d\right)\leq K_{7}\ln\left(t+j\right)\left(t^{-1}+t^{-d}j^{d-1}\right)$. For this, we take the derivative of the term after summation by parts and bound it in a similar way as in previous term.

Hence,

$$\begin{aligned}\frac{\partial}{\partial d}S_{t,j}&\simeq o\left(t^{1-d+\varepsilon}\right)\ln\left(t+j\right)\left[o\left(t^{-1}\right)+o\left(t^{-d}\right)o\left(j^{d-1}\right)\right]\\&=\ln\left(t+j\right)\left[o\left(t^{-d+\varepsilon}\right)+o\left(t^{1-2d+\varepsilon}\right)o\left(j^{d-1}\right)\right],\\\frac{\partial^{r}}{\partial d^{r}}S_{t,j}&\simeq o_{p}(T^{-r\tau})\ln^{r}\left(t+j\right)\left[o\left(t^{-d+\varepsilon}\right)+o\left(t^{1-2d+\varepsilon}\right)o\left(j^{d-1}\right)\right].\end{aligned}$$

Hence, after substituting in the Taylor approximation,

$$\begin{aligned} S_{t,j}(d) &= \left(d-\hat{d}\right)\ln(t+j)\left[o\left(t^{-d+\varepsilon}\right)+o\left(t^{1-2d+\varepsilon}\right)o\left(j^{d-1}\right)\right]+ \\ & \quad o\left(\ln(t+j)\,t^{-d+\varepsilon}\right)+o\left(\ln(t+j)\,t^{1-2d+\varepsilon}j^{d-1}\right) \text{ and} \\ \sum_{t=1}^{T} S_{t,j}(d) &\simeq o_p(T^{-\tau})\left[o\left(T^{1-d+\varepsilon}\right)+o\left(T^{2-2d+\varepsilon}\right)o\left(j^{d-1}\right)\right]. \end{aligned}$$

Finally,

$$\begin{aligned} \frac{1}{T^{3-2d}}\sum_{j=0}^{[\kappa T]} & \left(\sum_{t=1}^{T} S_{t,j}(d)\right)^2 = T^{2d-3}o_p(T^{-2\tau})\left[\left[o\left(T^{3-2d+\varepsilon}\right)\right]\right. \\ & \left. +o\left(T^{4-4d+\varepsilon}\right)\sum_{j=0}^{[\kappa T]} o\left(j^{2d-2}\right)\right] = o\left(T^{\varepsilon-2\tau}\right)=o(1)\,, \text{ for } \tau>0. \end{aligned}$$

Finally, for $0 < d < 1/2$, the proof follows from a similar argument.

Bibliography

[1] Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47-78.

[2] Beran, J., Terrin, N., 1996. Testing for a change of the long-memory parameter. Biometrika 83, 627-638.

[3] Boldea, O., Hall, A. R., 2010. Estimation and inference in unstable nonlinear least squares models. MPRA Paper 23150, University Library of Munich, Germany.

[4] Breitung, J. & Hassler, U., 2002. "Inference on the cointegration rank in fractionally integrated processes," Journal of Econometrics, Elsevier, vol. 110(2), pages 167-185, October.

[5] Brockwell, P.J., Davis, R.A., 1991. Time series: Theory and methods, second edition. Springer-Verlag, New York.

[6] Bunzel, H., Vogelsang, T. J., 2005. Powerful trend function tests that are robust to strong serial correlation, with an application to the Prebisch–Singer hypothesis. Journal of Business and Economic Statistics 23, 381-94.

[7] Busetti, F., Taylor, A. M. R., 2004. Tests of stationarity against a change in persistence. Journal of Econometrics 123, 33-66.

[8] Canjels, E., Watson, M.W., 1997. Estimating deterministic trends in the presence of serially correlated errors. The Review of Economics and Statistics 79, 184-200.

[9] Chung, C.F., Baillie, R.T., 1993. Small sample bias in conditional sum-of-squares estimators of fractionally integrated arma models. Empirical Economics 18, 791-806.

[10] Dahlhaus, R., 1995. Efficient location and regression estimation for long range dependent regression models. The Annals of Statistics 23, 1029-1047.

[11] Davidson, J., Hashimzade, N., 2009. Type I and type II fractional Brownian motions: a reconsideration. Computational Statistics and Data Analysis 53, 2089-2106.

[12] Deo, R.S., Hurvich, C.M., 1998. Linear trend with fractionally integrated rrrors. Journal of Time Series Analysis 19, 379-397.

[13] Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83-106.

[14] Dolado, J. & Gonzalo, J. & L. Mayoral, 2002. "A Fractional Dickey-Fuller Test for Unit Roots," Econometrica, Econometric Society, vol. 70(5), pages 1963-2006, September.

[15] Dolado, J.J., Gonzalo, J., Mayoral, L., 2005. What is what?: A simple time-domain test of long-memory vs. structural breaks. Economics Working Papers 954, Department of Economics and Business, Universitat Pompeu Fabra.

[16] .Dolado, J.J., Gonzalo, J., Mayoral L., 2008. Wald tests of I(1) against I(d) alternatives: some new properties and an extension to processes with trending components. Studies in Nonlinear Dynamics & Econometrics 12, 1-33.

[17] Dolado, J.J., Gonzalo, J., Mayoral, L. , 2009. Simple Wald tests of the fractional integration parameter: an overview of new results. In J. Castle & N. Shephard (eds.) The Methodology and Practice of Econometrics. A Festschift in Honour of David F. Hendry, 300-321, Oxford University Press.

[18] Durlauf, S.N., Phillips, P.C.B., 1988. Trends versus random walks in time series analysis. Econometrica 56, 1333-1354.

[19] Elliott, G., 1999. Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. International Economic Review 40, 767-783.

[20] Elliott, G., Mueller, U. K., 2006. Minimizing the impact of the initial condition on testing for unit roots. Journal of Econometrics 135, 285-310.

[21] Gadea, M. D., Mayoral, L., 2006. The persistence of inflation in OECD countries: a fractionally integrated approach. International Journal of Central Banking, 2, 51-104.

[22] Giacomini, R., Politis, D., White, H., 2007. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. Working Paper

[23] Gil-Alana, L. A., Robinson, P. M., 1997. Testing of unit root and other nonstationary hypotheses in macroeconomic time series. Journal of Econometrics 80, 241-268.

[24] Gil-Alana, L. A., 2008. Fractional integration and structural breaks at unknown periods of time. Journal of Time Series Analysis 29, 163-185.

[25] Gine, E., Zinn, J., 1990. Bootstrapping general empirical measures. Annals of Probability 18, 851-869.

[26] Giraitis, L., Kokoszka, P., Leipus, R., Teyssiere, G., 2003. Rescaled variance and related tests for long memory in volatility and levels. Journal of Econometrics 112, 265-294.

[27] Goodfriend, M., 2004. Inflation targeting in the united states? The Inflation-Targeting Debate, University of Chicago Press, http://www.nber.org/chapters/c9563.

[28] Granger, C. W. J., Hyung, N., 2004. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. Journal of Empirical Finance 11, 399-421.

[29] Granger, C. W. J. and Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis.1, 15–30.

[30] Grenander, U., Rosenblatt, M., 1957. Statistical analysis of stationary time series. Wiley, New York

[31] Hassler, U., Meller, B., 2011. Detecting multiple breaks in long memory: the case of US inflation. Discussion Paper Series 1: Economic Studies 2011,26, Deutsche Bundesbank, Research Centre.

[32] Hassler, U., Scheithauer, J., 2011. Detecting changes from short to long memory. Statistical Papers 52, 847-870.

[33] Hidalgo, J., Robinson, P.M., 1996. Testing for structural change in a long-memory environment. Journal of Econometrics 70, 159-174.

[34] Hosoya, Y., 2005. Fractional invariance principle. Journal of Time Series Analysis 26, 463-486.

[35] Hosking, J.R.M., 1996. Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. Journal of Econometrics 73, 261-284.

[36] Hualde, J., Robinson, P.M., 2011. Gaussian pseudo-maximum likelihood estimation of fractional time series models. Annals of Statistics 39, 3152-3181.

[37] Hsu, C.C., 2005. Long memory or structural changes: An empirical examination on inflation rates. Economics Letters 88, 289-294

[38] Johansen, S, Nielsen, M.Ø., 2010. Likelihood inference for a nonstationary fractional autoregressive model. Working Papers 1172, Queen's University, Department of Economics.

[39] Kapetanios, G., 2010. A generalization of a sieve bootstrap invariance principle to long memory processes. Quantitative and Qualitative Analysis in Social Sciences 4, 19-40.

[40] Kim, J.-Y., Belaire-Franch, J., Amador, R.B., 2002. Corrigendum to detection of change in persistence of a linear time series. Journal of Econometrics 109, 389-392.

[41] Kumar, M. S. and T. Okimoto, 2007. Dynamics of persistence in international inflation rates. Journal of Money, Credit and Banking 39, 1457–1479.

[42] Kuan, C.M. & Hsu, C.C., 1998. Change-point estimation of fractionally integrated processes. Journal of Time Series Analysis 19, 693-708.

[43] Lavielle, M., Ludeña C., 2000. The multiple change-points problem for the spectral distribution. Bernoulli 6, 845-869.

[44] Lavielle, M., Moulines E., 2000. Least squares estimation of an unknown number of shifts in a time series. Journal of Time Series Analysis 21, 33-59.

[45] Lavancier F., Leipus R., Philippe A., Surgailis D., 2011. Detection of non-constant long memory parameter. arXiv :1110.5138.

[46] Lazarova, S., 2005. Testing for structural change in regression with long memory processes. Journal of Econometrics 29, 329-372.

[47] Ling, S., Li, W.K., 2001. Asymptotic inference for nonstationary fractionally integrated autoregressive moving-average models. Econometric Theory 17, 738-764.

[48] Lobato, I. N., Savin, N. E., 1998. Real and spurious long-memory properties of stock-market data. Journal of Business & Economic Statistics 16, 261-268.

[49] Lobato, I. N., Velasco, C., 2007. Efficient Wald tests for fractional unit roots. Econometrica 75, 575-589.

[50] Lobato, I. N., Velasco, C., 2008. Power comparison among tests for fractional unit roots. Economics Letters 99, 152-154.

[51] Marinucci, D., Robinson, P.M., 1999. Alternative forms of fractional Brownian motion. Journal of Statistical Planning and Inference 80, 111-122

[52] Martins, L.F., Rodrigues, P.M.M., 2010. Testing for persistence change in fractionally integrated models: An application to world inflation rates. Working Papers w201030, Banco de Portugal, Economics and Research Department.

[53] Mayoral, L., 2006. Further evidence on the statistical properties of real gnp. Oxford Bulletin of Economics and Statistics 68, 901-920.

[54] Mayoral, L., 2012. Testing for fractional integration versus short memory with structural breaks. Oxford Bulletin of Economics and Statistics 74, 278-305.

[55] McCloskey, A., 2010. Semiparametric testing for changes in memory of otherwise stationary time series. Mimeo. January 2010.

[56] Michelacci, C., Zaffaroni, P., 2000. (Fractional) beta convergence. Journal of Monetary Economics 45, 129-153.

[57] Mikosch, T., Starica, C., 2004. Changes of structure in financial time series and the GARCH model. Econometrics 0412003, EconWPA.

[58] Mueller, U. K., Elliott, G., 2003. Tests for unit roots and the initial condition. Econometrica 71, 1269-86.

[59] Palma, W., 2007. Long-memory time series: Theory and methods. Wiley, J

[60] Perron, P., Zhu, X., 2005. Structural breaks with deterministic and stochastic trends. Journal of Econometrics 129, 65-119.

[61] Perron, P., Yabu, T., 2009. Estimating deterministic trends with an integrated or stationary noise component. Journal of Econometrics 151, 56-69.

[62] Perron, P., Qu, Z., 2010. Long-memory and level shifts in the volatility of stock market return indices. Journal of Business & Economic Statistics 28, 275-290.

[63] Pierre P. Zhou, J., 2008. Testing jointly for structural changes in the error variance and coefficients of a linear regression model. Working Papers Series wp2008-011, Boston University, Department of Economics.

[64] Phillips, P.C.B., Lee, C.C., 1996. Efficiency gains from quasi-differencing under nonstationarity. Cowles Foundation Discussion Papers 1134, Cowles Foundation, Yale University.

[65] Robinson, P. M., 1995. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630-1661.

[66] Shao, X., 2011. A simple test of changes in mean in the possible presence of long-range dependence. Journal of Time Series Analysis 32, 598-606.

[67] Shimotsu, K., 2006. Simple (but effective) tests of long memory versus structural breaks. Working Papers 1101, Queen's University, Department of Economics.

[68] Shimotsu, K., 2010. Exact local Whittle estimation of fractional integration with unknown mean and time trend. Econometric Theory 26, 501-540.

[69] Shimotsu, K., Phillips, P.C.B., 2005. Exact local Whittle estimation of fractional integration. Annals of Statistics 33, 1890-1933.

[70] Shimotsu, K., Phillips, P.C.B., 2006. Local Whittle estimation of fractional integration and some of its variants. Journal of Econometrics 130, 209-233.

[71] Sibbertsen, P., Kruse, R., 2009. Testing for a break in persistence under long-range dependencies. Journal of Time Series Analysis 30, 263-285.

[72] Sibbertsen, P., Willert, J., 2009. Testing for a break in persistence under long-range dependencies and mean shifts. Diskussionspapiere der Wirtschaftswissenschaftlichen Fakultät der Universität Hannover dp-422, Universität Hannover, Wirtschaftswissenschaftliche Fakultät.

[73] Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53, 165-188.

[74] Tanaka, K., 1999. The nonstationary fractional unit root. Econometric Theory 15, 549-582.

[75] Taylor, R., Harvey, D., Leybourne, S., 2010. Robust methods for detecting multiple level breaks in autocorrelated time series. Journal of Econometrics 157, 342-358.

[76] Velasco, C., 1999. Semiparametric gaussian estimation of non-stationary time series. Journal of Time Series Analysis 20, 87-127.

[77] Vogelsang, T. J., 1998. Trend function hypothesis testing in the presence of serial correlation. Econometrica 66, 123-48.

[78] Wright, J.H., 1995. Stochastic orders of magnitude associated with two-stage estimators of fractional arima systems. Journal of Time Series Analysis 16, 119–126.

[79] Wright, J. H., 1998. Testing for a structural break at unknown date with long-memory disturbances. Journal of Time Series Analysis 19, 369–376.

[80] Wooldridge, J. M., 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA. MIT Press.

[81] Yajima, Y., 1988. On estimation of a regression model with long-memory stationary errors. Annals of Statistics 16, 791-807.

[82] Yajima, Y., 1991. Asymptotic properties of the lse in a regression model with long memory stationary errors. Annals of Statistics 19, 158-177.

[83] Yamaguchi, K.,2011. Estimating a change point in the long memory parameter. Journal of Time Series Analysis 32, 304-314.